THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Atomic scale modeling of ordering phenomena

Mattias Ångqvist

Department of Physics
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2020

Atomic scale modeling of ordering phenomena
Mattias Ångqvist

© Mattias Ångqvist, 2020

isbn 978-91-7905-271-3

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr. 4738
issn 0346-718X

Department of Physics
Chalmers University of Technology
se-412 96 Göteborg, Sweden
Telephone +46 31 772 10 00

Cover: The icet logo.
Typeset in XeLaTeX.
Figures have been made in matplotlib [1], gnuplot [2], blender [3] and ASE [4].

Chalmers reproservice
Göteborg, Sweden 2020

Atomic scale modeling of ordering phenomena

Mattias Ångqvist
Department of Physics
Chalmers University of Technology

Abstract

Ordering phenomena in materials often have a crucial impact on materials properties. They are governed by the competition between entropy and energy. Accordingly simulating these aspects requires the construction of models that enable a computationally efficient exploration of the relevant configuration space. The alloy cluster expansion technique is particular well suited for this task as they can be trained to reach high accuracy while being computationally suitable for rapid sampling via Monte Carlo simulations.

In paper I we present the icet software for the construction and sampling of alloy cluster expansions. In this thesis the alloy cluster expansion method is applied to study several different materials.

The first group of materials studied are inorganic clathrates. In paper II and III we studied the ordering behavior and related properties as a function of composition and temperature for the clathrates $Ba_8Al_xSi_{46-x}$, $Ba_8Al_xGe_{46-x}$, $Ba_8Ga_xGe_{46-x}$ and $Ba_8Ga_xSi_{46-x}$. We achieved very good agreement with the available experimental data for the site occupancy factors (SOFs).

In paper IV and V we constructed the phase diagram for the W-Ti and W-C system respectively. A cluster expansion for each system was constructed and the configurational free energy was calculated. By also including other contributions to the free energy, most notably the vibrational free energy, the phase diagrams for these systems could be constructed.

In paper VI we studied the SSZ-13 zeolite and showed both that the Löwenstein rule is not respected with hydrogen as counterion and provided a rationale for this behavior.

Keywords: Cluster expansion, Monte Carlo, inorganic clathrates, alloys, zeolites, tungsten carbide, ordering phenomena

LIST OF PUBLICATIONS

This thesis consists of an introductory text and the following papers:

I **icet - A Python library for constructing and sampling alloy cluster expansions**
M. Ångqvist, W. A. Muñoz, J. M. Rahm, E. Fransson, C. Durniak, P. Rozyczko, T. H. Rod, and P. Erhart Advanced Simulation and Theory **2**, 1900015 (2019)

II **Optimization of the Thermoelectric Power Factor: Coupling between Chemical Order and Transport Properties**
Mattias Ångqvist, Daniel O. Lindroth and Paul Erhart
Chemistry of Materials **28**, 6877 (2016)

III **Understanding Chemical Ordering in Intermetallic Clathrates from Atomic Scale Simulations**
Mattias Ångqvist and Paul Erhart
Chemistry of Materials **29**, 7554 (2017)

IV **Structurally driven asymmetric miscibility in the phase diagram of W–Ti**
M. Ångqvist, J. M. Rahm, L. Gharaee, and P. Erhart Physical Review Materials 3, 30597 (2019)

V **Modeling of electronic, vibrational and configurational degrees of freedom in hexagonal and cubic tungsten carbide at high temperatures.**
Martin Gren, Erik Fransson, Mattias Ångqvist, Paul Erhart and Göran Wahnström (in manuscript)

VI **Generalization of the Löwenstein rule**
Magnus Fant, Mattias Ångqvist, Anders Hellman, Paul Erhart (in manuscript)

Specification of the authors contribution to the publications:

I The author wrote a large portion of the ICET code, wrote the paper along with the other authors, and carried out the simulations and analysis regarding the clathrates in the paper.

II The author wrote a draft for parts of the paper and constructed the cluster expansions, carried out the Monte Carlo simulations and associated analysis for the clathrate system. The author wrote the back-end parts of the code and major parts of the front-end.

III The author developed the software for constructing and sampling cluster expansions used in the paper. The author constructed the cluster expansions, carried out the Monte Carlo simulations and the associated analysis, and created all figures related to Monte Carlo simulations and cluster expansions.

IV The author developed the software for constructing and sampling cluster expansions used in the paper. The author constructed the cluster expansions and associated analysis, and prepared all figures cluster expansions.

V The author constructed the cluster expansions, carried out the Monte Carlo simulations and the associated analysis, and created all figures related to Monte Carlo simulations and cluster expansions, and wrote most of the configurational free energy section in the methods section.

VI The author developed the code needed for the Monte Carlo simulations, performed the Monte Carlo simulations, carried out the associated analysis, and made the figures related to cluster expansions.

Contents

1

Introduction

1.1 Atomic scale modeling of ordering phenomena

This thesis focuses on modeling ordering phenomena in materials using atomic scale models. It therefore relates to statistical and classical thermodynamics, computational physics, as well as materials physics. This introductory chapter introduces relevant thermodynamic aspects. The computational models used here are described in Chapters 2 and 3, whereas the studied materials are reviewed in chapter 4.

1.1.1 Entropy and order

Entropy is a measure of the disorder in a system. The second law of thermodynamics states that disorder in a closed system can only increase; in other words closed systems tend to evolve from ordered to disordered states. A simplified example of this is the process of shuffling a standard deck of cards. The initial state when the deck of cards are in perfect order will "never" appear again in the process of the shuffling since, assuming random shuffling, the probability to end up in the initial state is roughly 1 in 10^{68}. As the shuffling continues the deck of cards becomes more and more disordered.

Another example is the tossing of coins. Imagine having 100 coins and tossing them all at once. A particular sequence can be denoted as head–tail–head–head–… and so on. The probability that all come up heads is the probability of the first coin comes up as heads times the probability that the second coin comes up as heads and so on, hence the probability is $(1/2)^{100}$. This probability is the same for any sequence of the coin toss. Yet from intuition we would expect that all coins coming up as heads should almost never happen.

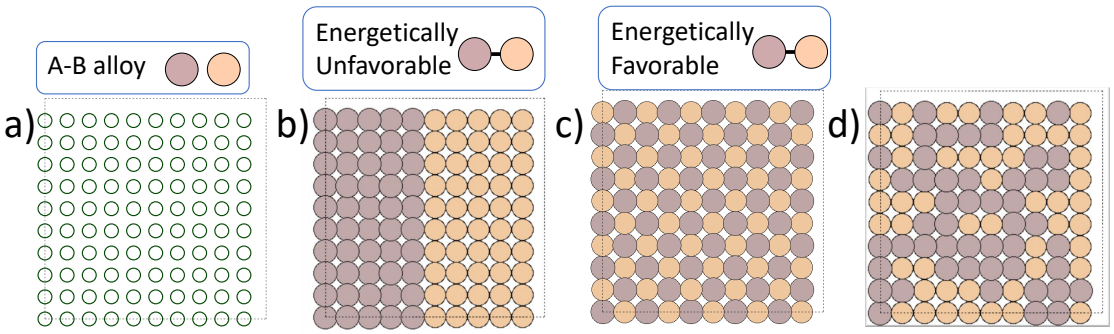

Figure 1.1: Chemical ordering it is essentially the issue of figure a), for a given A-B alloy at a certain temperature and composition what distribution of the atoms will be observed? Figure b), an ordered structure representing the ground state structure if A-B bonds are energetically unfavorable. Figure c), another ordered structure representing the ground state if A-B bonds are energetically favorable. Figure d), example of a disordered structure.

The resolution to this conundrum lies in the distinction of a specific sequence of coin tosses, called a microstate, and the total number of heads and tails of a particular sequence, called a macrostate. All microstates are equally probable but the probabilities for different macrostates vary over a wide range. The probability of a particular macrostate is the number of all microstates that correspond to that macrostate divided by the number of all possible microstates. In the coin-toss example, the number of total microstates is 2^{100} and the number of microstates for a macrostate is given by $\binom{100}{n} = 100!/n!(100-n)!$. Consequently the probability of all coins ending up as heads is $1/2^{100}$ whereas the probability of ending up with 50 heads is 10^{29} times more likely with a probability of about $10^{29}/2^{100}$. The number of microstates for a particular macrostate is commonly referred to as the multiplicity of that macrostate.

Mathematically entropy is measured as $S = k_B \ln \Omega$, where S is the entropy, k_B is the Boltzmann constant and Ω is the multiplicity of the system. Hence increasing the entropy, or the disorder, simply implies that the system changes to a more likely (macro)state.

The principles of the coin toss example are easily extended to atomic systems. Imagine a system comprised of 100 A atoms and as many lattice sites. By mixing one B atom into the system the number of possible microstates increases by a factor of 100, as there are 100 possible sites to insert the B atom. A system of atoms obeys the same combinatorics as coin flipping. Hence the entropy increases by mixing and if we are neglecting the atomic interactions the system will have a tendency to spontaneously start mixing.

1.1.2 Free energy

Entropy goes a long way in explaining the ordering behavior of materials. Yet, in nature many materials are found to exhibit ordered states, which have (much) lower entropy than disordered states. Thus if entropy is one half of the picture to explain ordering in materials, the energy is the other half that has to be included. Consider again the case of the AB atomic system, for which we saw that the entropy can be tremendously increased by mixing A and B atoms. In general there is an energetic cost associated with mixing. If the formation of $A-B$ bonds is energetically unfavorable compared to $A-A$ and $B-B$ bonds there is a penalty for mixing and less mixing is expected. On the other hand if $A-B$ bonds are more favorable, the system can both lower its energy and increase its entropy by mixing and more mixing is expected. This interplay of entropy and energy is expressed in the Helmholtz free energy of the system

$$F = U - TS, \tag{1.1}$$

where U is the internal energy and T is the temperature. Generally speaking a system described in the canonical ensemble will strive to minimize its free energy. According to Eq. (1.1) the entropy term becomes more important for higher temperatures. Hence, at a low temperature a system is more likely to be observed in its low energy state, where the chemical bonds dictate the ordering of the material. As the temperature is increased, however, the material becomes more disordered as the entropy term becomes more important.

1.1.3 Examples of ordering in materials

Many crystalline materials exhibit some form of chemical order which is crucial for understanding their properties. Here, one can distinguish between short and long-range order. As the terms suggest in the former case order only persists over a few neighbor shells, whereas long-range order persists over long, possibly macroscopic, distances.

Long-range order is commonly the result of strong attractive interactions and is associated with small periodic ordered structures. Short-range order on the other hand, is typically the result of a competition between energy and entropy as the attractive interactions are comparable in magnitude to the entropy. To model the latter type of systems it is therefore imperative to account for both of these contributions.

In many materials there is an order-disorder transition such that at temperatures $T < T_c$ the system is ordered and for temperatures $T > T_c$ the system is more or less disordered. The states below and above the transition can, e.g., be long-range and short-range ordered, respectively. One can intuitively guess that if a

material goes from an ordered state (Fig. 1.1 b) and c)) to a disordered state (Fig. 1.1 d)), the properties of the material can differ drastically. This behavior is particular pronounced and well studied in many alloys that form intermetallics, such as AuCu, FePt, or NiAl. It is also present in some of the materials studied in this thesis including the clathrates (section 4.1 and tungsten carbide (section 4.3).

Another type of ordering transition happens in Al based clathrates (see Paper III and section 4.1) where one can get a dramatic difference in chemical ordering for a small change in Al composition.

1.2 Mixing in multi-component systems

1.2.1 The dilute limit

Consider a system of N particles where all particles are of type A except for n particles of type B. This system is said to be in the dilute limit if $N \gg n$. In the dilute limit the B particles are so few and spread out that one can assume the interactions between B particles to be negligible. Hence, ΔF, the free energy associated with adding or subtracting one B particle, is independent of composition. The equilibrium concentration in the dilute limit is then given by $c = c_0 \exp\left[-\Delta F/k_B T\right]$, where c_0 is the concentration of sites available for substitution.

1.2.2 Concentrated solutions: Mean-field treatment

When the condition $N \gg n$ is no longer fulfilled the system is said to be in the concentrated limit. Here, the B atoms have become so numerous that their mutual interaction must be taken into account and the free energy of adding or removing a B atom becomes dependent on the concentration. The simplest treatment applied in the concentrated limit is provided by the Bragg-Williams model, which can be viewed as a mean field approximation to the Ising model. The energy, entropy, and the free energy of the system can be expressed in terms of an order parameter, which describes, e.g., the average number of $A-B$ bonds in the structure. The order parameter observed for a certain temperature and concentration is the one that minimizes F. The Helmholtz free energy of the system is given by [5, 6]

$$\begin{aligned} \Delta F_{\text{mix}} &= \Delta U_{\text{mix}} - T\Delta S_{\text{mix}}, \\ \Delta U_{\text{mix}} &= \omega c(1-c), \\ \Delta S_{\text{mix}} &= -k_B \left[c \ln c + (1-c)\ln(1-c)\right], \end{aligned} \tag{1.2}$$

where ω describes the energy cost of mixing and creating an $A-B$ bond, c is the B concentration and k_B is the Boltzmann constant. The effects of ΔU_{mix} and T on ΔF_{mix} are illustrated in Fig. 1.2. Some comments are in order. First, all expressions

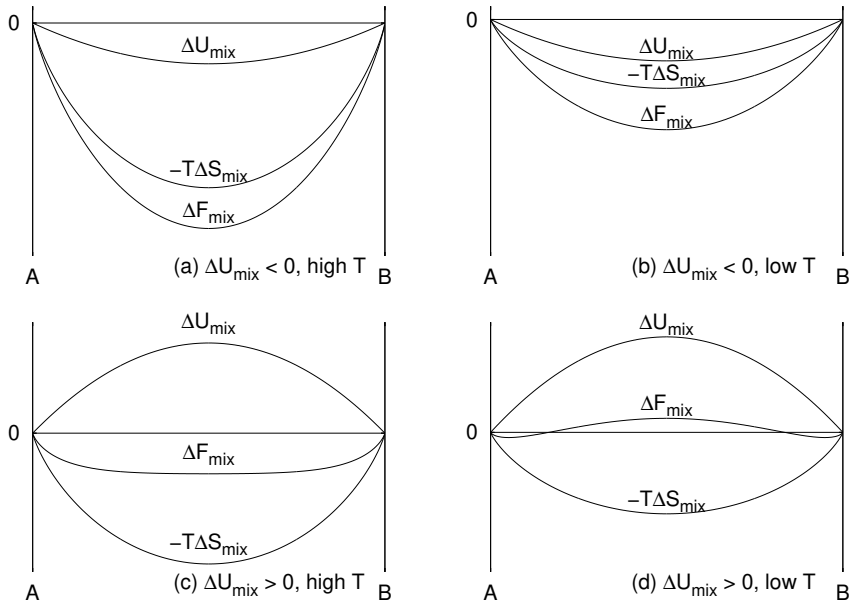

Figure 1.2: The free energy of mixing ΔF_{mix} as a function of composition assuming mixing is (a,b) energetically favorable ($\Delta U_{\text{mix}} < 0$) and (c,d) unfavorable ($\Delta U_{\text{mix}} > 0$), respectively, at (a,c) high and (b,d) low temperature T.

are symmetric in the concentration c. Secondly, the mixing entropy, ΔS, is positive across the entire concentration range. Consequently the contribution of entropy to the free energy is always negative. Hence, for $T \to \infty$, entropy will be the dominating term and the free energy will always be minimized by mixing ($\Delta F_{\text{mix}} < 0$). For the case of $\Delta U_{\text{mix}} > 0$ there is a competition between energy and entropy that determines whether the components mix or segregate. For $\Delta U_{\text{mix}} > 0$ and low T, ΔF_{mix} can change sign, giving rise to a miscibility gap (Fig. 1.2d).

This very simple form is not suitable for modeling the vast majority of systems. It can, however, be generalized leading to the semi-empirical CALPHAD approach to modeling phase diagrams. In this case, the various contributions to the mixing energy and entropy are expressed in polynomial expansions in temperature and composition. The expansion coefficients are most commonly obtained by fitting to experimental data and, more recently, also from first-principles calculations. The resulting models are widely used in industry and academia for alloy design and optimization.

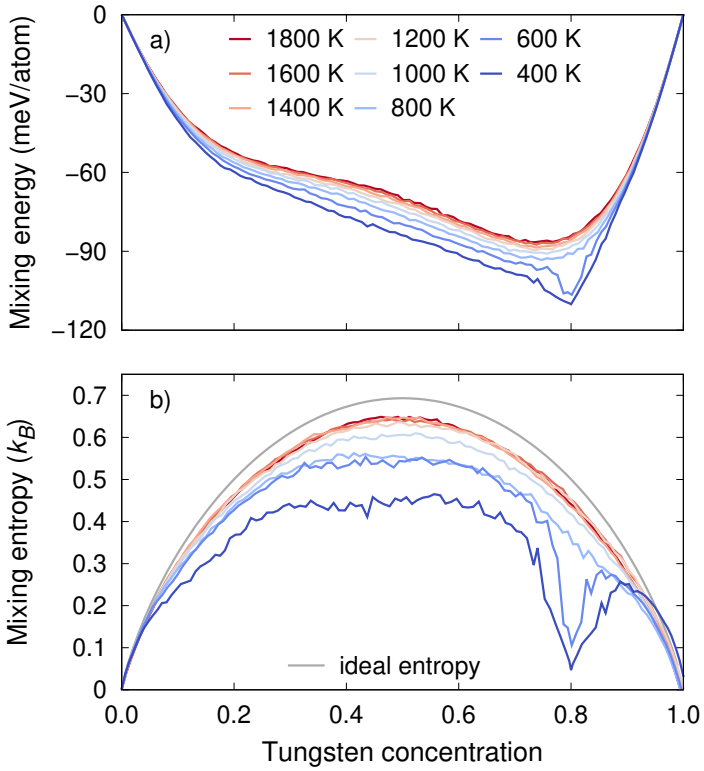

Figure 1.3: (a) Mixing energy and (b) mixing entropy as a function of composition.

1.2.3 Concentrated solutions: Beyond mean-field

The mean-field approach outlined in the previous section is fundamentally semi-empirical in its nature and requires input in the form of either experimental or computational data. Experimental data is often difficult and/or expensive to acquire. There are also various cases, in which it is downright impossible to extract meaningful data from experiments. This applies for example at low temperatures when thermodynamic equilibrium cannot be reached reliably.

For illustration consider the W–Ti system as analyzed using a combination of first-principles calculations, lattice models and Monte Carlo simulations (Paper IV). The analysis reveals a ground state structure at 80% tungsten concentration [7]. At low temperatures this configuration is much more likely to be observed than other configurations leading to a distinct feature in the mixing energy (Fig. 1.3a) and even more so the mixing entropy (Fig. 1.3b). These features are absent in CALPHAD assessments of the W–Ti system, which instead commonly assume complete immiscibility, i.e. a positive mixing energy throughout. This approximation is owed to the fact that W is a refractory metal, which renders experimental data below approximately 1300 K unreliable.

1.3 Thesis guide

The W–Ti system alluded to in the previous section is one example for how the combination of first-principles calculations (usually density functional theory (DFT)) and lattice models (usually alloy cluster expansions (CEs)) allows one to accurately predict the behavior of multi-component mixtures with little or no experimental input.

This thesis comprises the development of a computational tool for constructing and sampling CEs (Paper I). This is the approach taken in this thesis, which in addition to the W–Ti alloy system (Paper IV), extends the application of these techniques to the analysis of chemical ordering in inorganic clathrates (Papers II and III) and zeolites (paper VI) as well as vacancy ordering in tungsten carbides (paper V).

Alloy CEs are introduced in chapter 2. This method is a central part of this thesis and used in all papers. The sampling of CEs in different thermodynamic ensembles is covered in chapter 3. In this thesis ordering phenomena has been studied in several different materials. In chapter 4 these materials are introduced, their importance and the role ordering is discussed. Finally a summary of the papers included in this thesis can be found in chapter 5.

2

Alloy cluster expansions

2.1 Introduction

The partition function $\mathcal{Z}$, contains all thermodynamic information of a system To compute $\mathcal{Z}$ one needs to calculate the potential energy for each possible microstate of the system. For this to be feasible, however, one needs both efficient and accurate energy calculations, which can be achieved using the alloy CE technique. In the CE formalism the system of interest is described by an occupation vector $\vec{\sigma}$ where σ_i can, for a binary system, assume a value of either 0 or 1 depending on if an A or B atom is on lattice point i (Figure 2.1). Even though the cluster expansion acts on a perfect lattice it can still capture the contribution of relaxations of the atoms by using the energies of relaxed structures for constructing the model.

2.2 Definition of a cluster

A cluster is defined as a set of lattice points, $\vec{\alpha} = \{\sigma_1, \sigma_2, ..., \sigma_n\}$. A cluster is thus associated with a structure and possibly periodic boundary conditions. The order of a cluster is defined as the number of lattice points in the cluster. A cluster of order 1 is called a singlet and clusters of order 2, 3 and 4 are referred to as pairs, triplets, and quadruplets, respectively. The radius, or size, of the cluster can be defined as the average distance of all the lattice points from the geometric center of the cluster. For a given lattice a set of clusters can be conveniently defined as a set of cutoffs, $\vec{r}^{\text{cutoff}} = \{r_{\text{pair}}, r_{\text{triplet}}, ...\}$, such that only pair (triplet) clusters with a maximum interatomic distance of less than r_{pair} (r_{triplet}) will be included. For illustration, Figure 2.3 shows the clusters in the body-centered cubic (BCC) lattice up to sixth order.

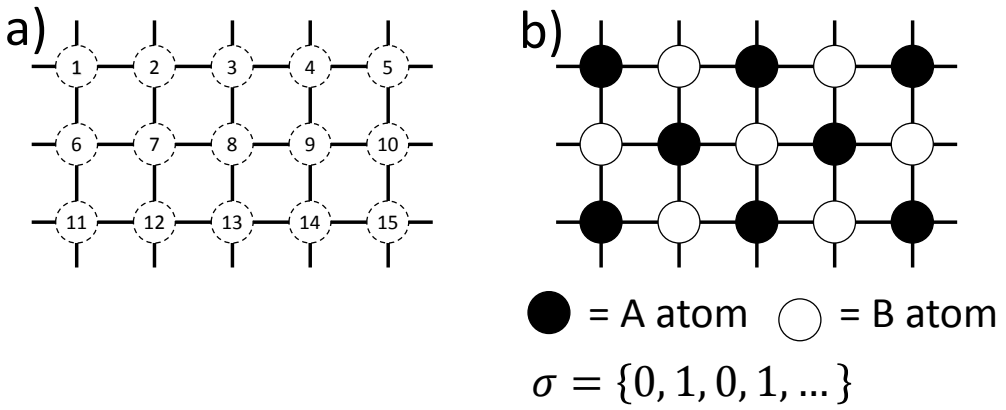

Figure 2.1: a) A fixed lattice with numbered lattice points. b) When the lattice points are occupied by atoms the state of the system can be described by the $\vec{\sigma}$ vector where σ_i is the occupation on site i.

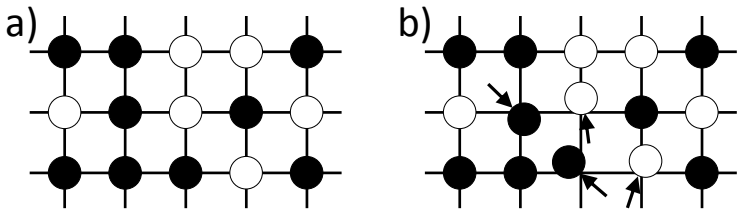

Figure 2.2: Two microstates of the lattice with the same $\vec{\sigma}$ vector. a) The atoms sit perfectly on the underlying lattice. b) Some atoms have small displacements away from the ideal positions.

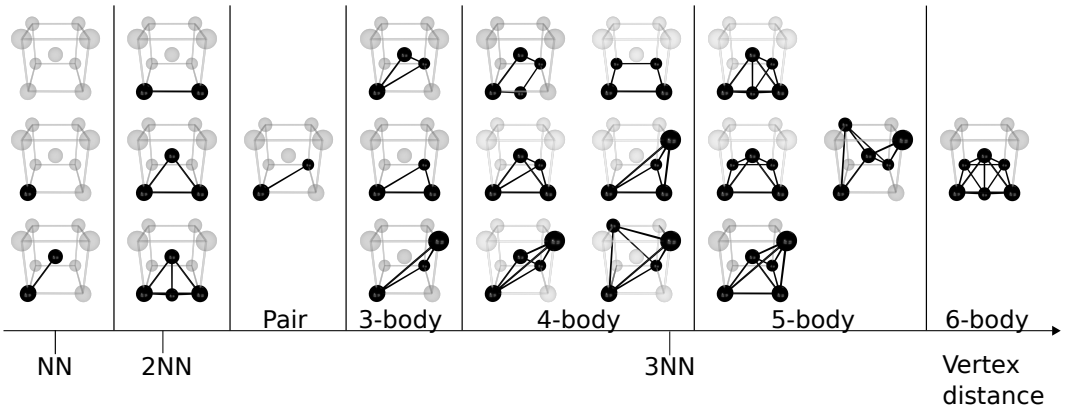

Figure 2.3: Illustration of the smallest clusters up to sixth order in a BCC lattice.

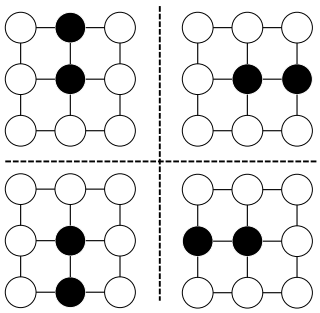

Figure 2.4: Symmetrically equivalent first nearest neighbor pair clusters in a square lattice. These can be found by taking any one of the specific decorations and repeatedly applying a 90 degree rotation until all four equivalent clusters are found.

2.3 Formal theory

A CE is able to represent *any* function of the configuration, $f(\vec{\sigma})$ if one can construct a complete orthogonal basis of functions with respect to the scalar product [8]

$$\langle f, g \rangle = \frac{1}{M^N} \sum_{\vec{\sigma}_1 \ldots, \vec{\sigma}_{M^N}} f(\sigma) g(\sigma)), \tag{2.1}$$

where $f(\vec{\sigma})$ and $g(\vec{\sigma})$ are two arbitrary functions of the configuration, M is the allowed number of elements, and N is the number of lattice points in $\vec{\sigma}$.

2.3.1 Point functions

For each lattice point p, we define the M orthogonal point functions $\Theta_n(\sigma_p)$

$$\Theta_n(\sigma_p) = \begin{cases} 1 & \text{if } n = 0 \\ -\cos\left(\pi(n+1)\sigma_p/M\right) & \text{if } n \text{ is odd} \\ -\sin\left(\pi n \sigma_p/M\right) & \text{if } n \text{ is even.} \end{cases} \tag{2.2}$$

It can be verified that these point functions form an orthogonal set over all possible occupation numbers [9],

$$\langle \Theta_n, \Theta_{n'} \rangle = \sum_{\sigma_p=0}^{M-1} \Theta_n(\sigma_p)\Theta_{n'}(\sigma_p) = \begin{cases} 0 & \text{if } n \neq n' \\ \neq 0 & \text{if } n = n'. \end{cases} \tag{2.3}$$

For example, in a three component system ($M = 3$) with $\sigma_p = \{0, 1, 2\}$ the possible point functions are

$$\Theta_0(\sigma_i) = 1, \quad \Theta_1(\sigma_i) = -\cos 2\pi\frac{\sigma_i}{3}, \quad \Theta_2(\sigma_i) = -\sin 2\pi\frac{\sigma_i}{3}.. \tag{2.4}$$

2.3.2 Orthogonal basis

Using these point functions an orthogonal, one can produce a set of functions $\Pi_\alpha^{(s)}(\vec{\sigma})$ in the space of the M^N configurations on the lattice by generating the point functions for all possible combinations of s and lattice points $\vec{\alpha}$. So for a cluster of lattice sites $\vec{\alpha} = \{1, 2, ..., |\vec{\alpha}|\}$, and a vector of allowed point function indices, $s = n_1, n_2, ..., n_l$ the basis functions are given by

$$\Pi_\alpha^{(s)}(\sigma) = \Theta_{n_1}(\sigma_1)\Theta_{n_2}(\sigma_2)...\Theta_{n_l}(\sigma_\alpha) \tag{2.5}$$

and it can be verified that these form an orthogonal set[9],

$$\left\langle \Pi_\alpha^{(s)}, \Pi_\beta^{(s')} \right\rangle = \delta_{\alpha\beta}\delta_{ss'}. \tag{2.6}$$

Since the basis functions $\Pi_\alpha^{(s)}$ form an orthogonal set we can express any function of the configuration as

$$f(\sigma) = \sum_\alpha \sum_s f_{\alpha s}\Pi_\alpha^{(s)}(\sigma). \tag{2.7}$$

Since all basis functions $\Pi_\alpha^{(s)}$ have one configuration invariant component that is equal to 1 when $s = \{0, 0, ..., 0\}$ we can exclude this term from the sum in Eq. (2.7) to obtain

$$f(\sigma) = f_0 + \sum_\alpha \sum_s f_{\alpha s}\Pi_\alpha^{(s)}(\sigma). \tag{2.8}$$

Finally, for practical reasons, we modify Eq. (2.8) by averaging over each distinct cluster and point functions and multiply with the multiplicity to arrive at the final expression for our cluster expansion function

$$f(\sigma) = f_0 + \sum_\alpha \sum_s \left\langle \Pi_\alpha^{(s)}(\sigma) \right\rangle_{\alpha'} m_\alpha^{(s)} J_\alpha^{(s)}. \tag{2.9}$$

Here, the summation is carried out over all symmetrically distinct clusters of lattice points. The $\langle ... \rangle_{\alpha'}$ function takes the average over the basis functions for all clusters α' that are symmetry equivalent to α. $J_\alpha^{(s)}$ are the effective cluster interactions (ECIs), which determine a specific cluster expansion. Finally, $m_\alpha^{(s)}$ is the multiplicity of cluster α for a specific combinations of point functions $\vec{s}$. Note that some authors choose to include $m_\alpha^{(s)}$ in $J_\alpha^{(s)}$.

2.3.3 Further considerations

As we have seen from the construction of the basis, all combinations of point functions are required for constructing the basis. For a binary system where only the first point function was needed the permutations of these point functions for any cluster order are all equal. For a ternary system both the first and second point functions are needed and thus a pair will have four different combinations of the point functions, i.e. (1, 1), (1, 2), (2, 1) and (2, 2). As will be shown now, not all of these permutations will result in additional parameters in the CE due to symmetry. If our pair cluster $\alpha = \{\sigma_1, \sigma_2\}$ can be transformed under periodic boundary conditions and the lattice translational and rotational symmetry operations to construct the symmetrically equivalent cluster $\alpha' = \{\sigma_2, \sigma_1\}$ then the choice of the ordering in $\alpha = \{\sigma_1, \sigma_2\}$ must produce the same contribution to Eq. (2.9) as choosing $\alpha = \{\sigma_2, \sigma_1\}$. Writing out the different choices we have for α

$$\begin{aligned}
\Pi_{\alpha}^{1,1}(\sigma) &= \Theta_1(\sigma_1)\Theta_1(\sigma_2) \\
\Pi_{\alpha}^{1,2}(\sigma) &= \Theta_1(\sigma_1)\Theta_2(\sigma_2) \\
\Pi_{\alpha}^{2,1}(\sigma) &= \Theta_2(\sigma_1)\Theta_1(\sigma_2) \\
\Pi_{\alpha}^{2,2}(\sigma) &= \Theta_2(\sigma_1)\Theta_2(\sigma_2)
\end{aligned} \tag{2.10}$$

and for α'

$$\begin{aligned}
\Pi_{\alpha'}^{1,1}(\sigma) &= \Theta_1(\sigma_2)\Theta_1(\sigma_1) \\
\Pi_{\alpha'}^{1,2}(\sigma) &= \Theta_1(\sigma_2)\Theta_2(\sigma_1) \\
\Pi_{\alpha'}^{2,1}(\sigma) &= \Theta_2(\sigma_2)\Theta_1(\sigma_1) \\
\Pi_{\alpha'}^{2,2}(\sigma) &= \Theta_2(\sigma_2)\Theta_2(\sigma_1).
\end{aligned} \tag{2.11}$$

The basis functions (1, 1) and (2, 2) are thus symmetric in terms of permuting the lattice points. For (1, 2) and (2, 1) there is, however, an asymmetry and the choice of the order in the cluster α will matter for the end result in Eq. (2.9). The choice of ordering of the lattice points in a cluster is completely arbitrary and should not matter to the final result of the cluster expansion. To circumvent the choice of ordering of lattice points in a cluster we instead only use the point functions (1, 1), (1, 2) and (2, 2) for this cluster. Additionally, for the cluster basis (1, 2) we use both possibilities of ordering of the pair. The final basis functions for cluster α then become the following

$$\begin{aligned}
\Pi_{\alpha}^{1,1}(\sigma) &= \Theta_1(\sigma_1)\Theta_1(\sigma_2) \\
\Pi_{\alpha}^{1,2}(\sigma) &= \Theta_1(\sigma_1)\Theta_2(\sigma_2) + \Theta_1(\sigma_2)\Theta_2(\sigma_1), \\
\Pi_{\alpha}^{2,2}(\sigma) &= \Theta_2(\sigma_1)\Theta_2(\sigma_2)
\end{aligned} \tag{2.12}$$

where it it is apparent that the choice of the order in the cluster has no effect on its representation in the cluster space. Note that in Eq. (2.9) the multiplicity $m_{\alpha}^{(1,2)}$ will be twice as large as the other multiplicities.

2.3.4 Symmetrically indistinct clusters

This section describes the identification of equivalent clusters by using symmetry operations. For the n-body cluster $\alpha = \{\sigma_0, \sigma_1, .., \sigma_{n-1}\}$ the symmetrically equivalent clusters are found by converting the lattice points into fractional positions $\alpha = \{\vec{r}_0, \vec{r}_1, .., \vec{r}_{n-1}\}$. A symmetry operation $\hat{\vec{s}}$ consists of a linear transformation by a 3×3 matrix $\bar{\gamma}$ and a translation $\vec{\tau}$

$$\vec{r}\,' = \hat{\vec{s}}\vec{r} = \bar{\gamma}\vec{r} + \vec{\tau}. \tag{2.13}$$

A lattice typically has a number of associated symmetry operations $\vec{S} = \hat{\vec{s}}_0$, $\hat{\vec{s}}_1$, ..., $\hat{\vec{s}}_{m-1}$. These symmetry operations can then be used to produce m symmetrically indistinct clusters where the cluster i is given by

$$\alpha_i = \{\hat{\vec{s}}_i\vec{r}_0, \hat{\vec{s}}_i\vec{r}_1, .., \hat{\vec{s}}_i\vec{r}_{n-1}\}. \tag{2.14}$$

Additionally, depending on the periodic boundary conditions, one can also translate all positions in a cluster with multiples of the unit cell vectors. Figure 2.4 illustrates the set of equivalent clusters for the nearest neighbor pair.

2.4 Construction of a cluster expansion

Equation (2.9) can represent any function of the configuration $f(\vec{\sigma})$. The task remaining in the construction of a CE is to find the ECIs that appear in Eq. (2.9). To this end, one requires reference data in the form of a set of configurations $\{\vec{\sigma_1}, \vec{\sigma_2}, ..., \vec{\sigma_n}\}$ as well as target data $\{E_1, E_2, ..., E_n\}$. The sums in Eq. (2.9) can be replaced with a dot product

$$f(\vec{\sigma}) = f_0 + \sum_{\alpha} \sum_{s} \left\langle \Pi_{\alpha}^{(s)}(\sigma) \right\rangle_{\alpha'} m_{\alpha}^{(s)} J_{\alpha}^{(s)} = \vec{\omega(\vec{\sigma})} \cdot \vec{J}, \tag{2.15}$$

where

$$\vec{\omega(\vec{\sigma})} = \Bigg\{ 1, \left\langle \Pi_{\alpha_1}^{(s_\alpha)}(\sigma) \right\rangle_{\alpha_1'} m_{\alpha_1}^{(s_{\alpha_1})}, ... \left\langle \Pi_{\alpha_1}^{(s'_{\alpha_1})}(\sigma) \right\rangle_{\alpha_1'} m_{\alpha_1}^{(s'_{\alpha_1})}, \\ ... \left\langle \Pi_{\alpha_n}^{(s_{\alpha_n})}(\sigma) \right\rangle_{\alpha_n'} m_{\alpha_n}^{(s_{\alpha_n})} \Bigg\},$$

and $\vec{J}$ denotes the vector of ECIs where $J_0 = f_0$. Here, we refer to the vector $\vec{\omega(\vec{\sigma})}$ as the cluster vector. Note that it can sometimes be useful to move $m_{\alpha}^{(s)}$ from $\vec{\omega}$ into $\vec{J}$ and let the target values E_i refer to the primitive unit cell. This will ensure all elements in $\vec{\omega}$ are in the interval $[-1, 1]$ and avoid a bias due to the number of

elements in $\vec{\sigma}$. Now we can cast the problem of finding the ECIs in the form of a linear equation

$$\bar{\Pi}\vec{J} = \vec{E}, \tag{2.16}$$

with

$$\bar{\Pi}^\top = [\vec{\omega}(\vec{\sigma_1}), \vec{\omega}(\vec{\sigma_2}), ... \vec{\omega}(\vec{\sigma_n})]\,, \tag{2.17}$$

$$\vec{J}^\top = \left[J^{(s)}_{\alpha_1}, J^{(s)}_{\alpha_2}, ... J^{(s)}_{\alpha_n}\right], \tag{2.18}$$

$$\vec{E}^\top = [E_1, E_2, ... E_n]\,. \tag{2.19}$$

The ECIs $J^{(s)}_\alpha$ are unknown and since the number of clusters is in principle infinite there is an infinite number of unknown parameters to determine. Based on physical intuition we expect that physical interactions are short-ranged and few-bodied. Therefore, if we construct our basis functions starting from singlets and geometrically small pairs, triplets etc. the CE is expected to converge quickly to yield an acceptable tolerance and the number of unknown parameters remains manageable.

2.5 Linear regression techniques

2.5.1 Ordinary least squares

There are several ways to solve Eq. (2.19) and find the ECIs. One of the simplest ways is to use ordinary least-squares (OLS), i.e. to find the $\vec{J}$ that minimizes $\left\|\vec{E} - \bar{\Pi}\vec{J}\right\|_2$, where $\bar{\Pi}$ is the matrix of cluster vectors, $\vec{J}$ is the vector of ECIs and $\vec{E}$ is the vector of target properties. The disadvantage of OLS is that it requires the problem to be over-determined, which quickly becomes computationally expensive as the training data typically comes from DFT calculations. Another downside is that the solutions provided by OLS are commonly dense, i.e. all elements of $\vec{J}$ have non-zero values, which can be an issue for the final model in terms of computational effort. OLS is therefore prone to overfitting, which will substantially deteriorate the ability of the model to predict physical properties [10].

2.5.2 Regularization and over-fitting

To overcome the disadvantages of bare OLS one can add regularization terms to the objective function and/or employ explicit feature selection algorithms such as recursive feature elimination (RFE). To illustrate the approach consider the expression for so-called elastic net optimization

$$\vec{J} = \arg\min_J \left\{ \alpha \left\|\vec{J}\right\|_1 + \beta \left\|\vec{J}\right\|_2^2 + \frac{1}{2}\left\|\vec{E} - \bar{\Pi}\vec{J}\right\|^2 \right\}, \tag{2.20}$$

where α and β are so-called hyperparameters, i.e. parameters that are associated with the optimization algorithm as opposed to the model itself. For $\alpha = 0$ one obtains ridge regression while for $\beta = 0$ one recovers the least absolute shrinkage and selection operator (LASSO). The purpose of the terms associated with α and β is to force the solution vector $\vec{J}$ to be "short" and combat overfitting.

For illustration consider the LASSO expression

$$\vec{J} = \arg\min_{J} \left\{ \alpha \left\| \vec{J} \right\|_1 + \frac{1}{2} \left\| \vec{E} - \bar{\Pi}\vec{J} \right\|^2 \right\}. \tag{2.21}$$

A high α parameter punishes large l_1 norms of the solution vector while ensuring that the original OLS problem is still solved. In theory this allows one to construct a large cluster basis and extract the relevant ECIs with just a few training structures. In other words, this approach in principle enables one to obtain meaningful solutions for heavily underdetermined systems. By ensuring the vector $\vec{J}$ is sparse this approach is expected to ensure "physicality" of the solution in the sense of interactions being short-ranged and few-bodied. In practice, however, it is often not that simple.

Regularization for CE construction has been demonstrated for the first time in 2013 [11, 12, 13] and has been very successful since. It was introduced under the name of compressive sensing, which is a task in signal recovery that is usually approached by finding solutions to an underdetermined linear system but does *not* itself solve the linear problem. Rather the compressive sensing field has provided theorems for the existence of solutions and the probability for finding correct solutions [14]. Unfortunately, this occasionally leads to a confusion of terminology.

The original papers [11, 12] employed the split-Bregman algorithm, which is one practical approach to solving the LASSO problem. There are, however, various other techniques available for approaching the general problem of solving both over and underdetermined linear problems, including LASSO and variants thereof such as adaptive LASSO [15] or group LASSO [16], elastic net, ridge regression, automatic relevance determination regression (ARDR) [17] or RFE.

2.5.3 Recursive feature selection

Among the aforementioned techniques, RFE plays a special role as it is not a minimization algorithm but a pure feature selection approach (several of the other approaches combine these aspects). In RFE the underlying minimization algorithm is used to train a model, after which the weakest parameter(s) are removed until a specified number of features n_f is reached. The target n_f is the hyperparameter associated with RFE.

RFE can thus be combined with any of the minimization techniques, although RFE with OLS is the common. The latter combination is computationally very efficient but has the downside of only working in the overdetermined limit.

2.5.4 Practical considerations

In practice an optimization procedure should (1) be computationally efficient, (2) require little to no adjustments, and (3) produce physically sound sparse models.

Computational efficiency. OLS is computationally the most efficient technique. Since all other approaches in some form or another require solving the OLS problem multiple times during runtime, they are necessarily (many times) more expensive than standard OLS. An analysis of the scaling behavior of these methods in the context of force constant expansions shows that the computational cost only becomes a significant for models with several thousand target values [10]. In practice computational efficiency is therefore seldom a limit for CE construction.

Ease of use. As most of the algorithms mentioned above are available in general purpose implementations, the main consideration here is the number of hyperparameters that require adjustment and the sensitivity of the final model to these parameters. In the case of LASSO and ARDR there is effectively one hyperparameter that requires adjustment (although there are internally more). Split-Bregman requires two parameters to be adjusted. RFE adds another hyperparameter to any method it is coupled to. In practice, optimal hyperparameters can be determined automatically, usually in combination with cross-validation (CV) (see below). As a result all of the methods here are comparable in terms of their ease of use.

Physically sound sparse models. Among the aforementioned techniques, LASSO, ARDR, and any method in combination with RFE yield solution vectors $\vec{J}$ with elements that are strictly zero or non-zero. This is very useful from a computational perspective as it allows one to include a much smaller number of ECIs when sampling the CE with e.g., Monte Carlo (MC) simulations. In the case of split-Bregman additional steps are required to remove ECIs that are statistically insignificant. With respect to the quality of the models, it has been found that ARDR and RFE-OLS (used in the later papers IV, V, VI) outperform LASSO or split-Bregman (used in the earlier papers II and III). While these insights evolved over time by practical experience, the behavior of the different algorithms was then more systematically investigated in Paper I, which shows that in particular ARDR has notable advantages. It works well both in the over and underdetermined limits (unlike RFE-OLS) as well as the transition region (unlike e.g., LASSO). It

consistently generates sparse solutions (unlike LASSO or split-Bregman with in particular the former yielding a lot of false positives). Finally, it converges reliably and is overall very stable. A more comprehensive analysis of these algorithms in the context of force constant expansions [10] further supports these findings. It also shows ARDR to exhibit an unfavorable scaling with the size of the training set, which is, however, only relevant for training sets with several thousand entries and hence usually of no concern for CE construction.

Practical implementation. Many of the techniques described above are available via SCIKIT-LEARN [18, 19], a Python library that allows easy access to several advanced machine learning linear regression techniques. The ICET package (Paper I) is interfaced to SCIKIT-LEARN making it easy and user friendly to test different algorithms for training CEs.

2.6 Cross-validation

Commonly when training models the available data is split into training and test sets where the training set is used for optimization and the test set for evaluating the accuracy. The usefulness of these different sets and errors become apparent when trying to select a good model. For example, if for a particular hyperparameter, e.g. the α for LASSO, the training error is zero but the test error is high, the trained model is excellent at predicting already seen data but unable to predict new and unseen data. This is called overfitting and must be avoided.

Another type of overfitting is encountered when one instead finds optimal values of hyperparameters that yield the minimum test error. Even though the test set is not part of the training set, information about the training set still seeps into the training procedure since the parameters obtained in the training are optimal for the specific test set. To overcome this slightly more hidden type of overfitting one typically carries out many rounds of training and test splits in a procedure known as cross-validation.

2.6.1 Leave-one-out cross validation

One type of CV is the leave-one-out CV (LOO-CV), which is defined as

$$(\mathrm{CV})^2 = \frac{1}{N}\sum_{n=1}^{N}\left(\hat{E}_{(n)} - E_n\right)^2, \tag{2.22}$$

where E_n is the calculated energy for structure n and $\hat{E}_{(n)}$ is the predicted value of the energy of structure n as calculated with the CE trained with the $(N-1)$ other

structures. While this method appears to require making N cluster expansions, one can analytically show that it is equivalent to training one CE with the full set, after which one can compute the LOO-CV score by rescaling the root-mean-square error (RMSE) over the training set using the correlation of the fit matrix.

When the training curve is steep and the number of available structures is low the LOO-CV can be a good estimator.

2.6.2 k-fold cross validation

Another commonly used estimator for the error is k-fold CV. Here, the data is randomly divided up into k evenly sized subsets. One of the k subsets is left out for validation and the remaining $k-1$ subsets are used for training. This is repeated for all k subsets and the final cross validation score is the average over the k splits. Note that k-fold validation reduces to LOO-CV when k equals the number of available data points. Sometimes it can be desirable to improve statistics of the CV score beyond what is obtained from a single k-fold cross validation estimation. This can happen if one wants to estimate the CV score with a training size of two thirds resulting in only three folds. The options are then to either repeat the k-fold procedure and pick another choice of the sets or do shuffle and split.

2.6.3 Shuffle-and-split

The CV score can also be estimated with a technique known as shuffle-and-split. Given a training set size, the training set is obtained by randomly drawing structures from the entire available set and then testing against the remaining structures. Here, it is common to select with replacement, an approach referred to as bagging. The procedure can then be repeated as many times as needed to get a reliable estimate of the CV score. One downside of this method is that several training structures can occur in the test or train set in a row due to the pseudo-random nature of the algorithm which might lead to a selection bias. The upside, however, is that this method offers an easier way to do repeated training-test splits than the k-fold method.

2.7 Cluster expansion for a binary system

To demonstrate the formalism developed in this chapter we will now explicitly describe the construction of a CE for simple binary system. In a binary system only one point function will be used in the cluster functions, $\Theta_1(\sigma_p) = -\cos\left(\pi\sigma_p\right)$. Only two occupation numbers are needed as well, 0 and 1, which have corresponding point function values of -1 and $+1$, respectively. White and black atoms are to

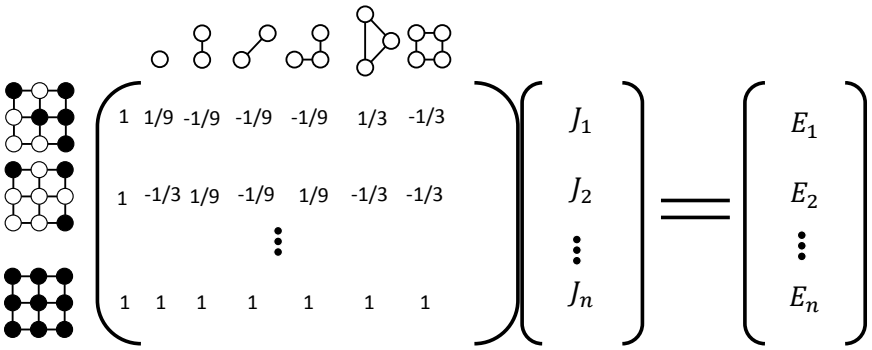

Figure 2.5: Cluster expansions can be constructed by solving a set of linear equation systems.

be indicated by occupation 0 and 1, respectively. The average over symmetrically distinct clusters α' in Eq. (2.9) for the singlet cluster is

$$\left\langle \Pi_{\alpha=\text{singlet}}(\vec{\sigma}) \right\rangle_{\alpha'} = \frac{N_{\vec{\sigma}_{\alpha'}=(1)} - N_{\vec{\sigma}_{\alpha'}=(0)}}{N_{\vec{\sigma}_{\alpha'}=(0\,\|\,1)}}. \tag{2.23}$$

For a pair the averages become

$$\left\langle \Pi_{\alpha=\text{pair}}(\vec{\sigma}) \right\rangle_{\alpha'} = \frac{N_{\vec{\sigma}_{\alpha'}=(0,0)} + N_{\vec{\sigma}_{\alpha'}=(1,1)} - N_{\vec{\sigma}_{\alpha'}=(0,1)}}{N_{\vec{\sigma}_{\alpha'}=(0,0\,\|\,0,1\,\|\,1,1)}}, \tag{2.24}$$

where $N_{\vec{\sigma}_{\alpha'}=(i,j)}$ is the number of equivalent clusters that have occupation (i,j) or (j,i). To calculate values such as $N_{\alpha'=(0,0)}$ a summation is done over all index pairs (i,j) that are symmetrically equivalent. Hence, the cluster vector $\vec{\omega}(\vec{\sigma})$ is obtained by summation and averaging of the cluster functions. All the $\vec{\alpha}$ can be precomputed so a summation can be performed very efficiently. Then all available configurations can be mapped to cluster vectors $\vec{\omega}$, representing training and test sets (Fig. 2.5) as described in the previous section. Once suitable values for the hyperparameters have been found, by using various cross validation techniques, the final set of ECIs, $\vec{J}_{\text{final}}$, can be obtained. The construction of the cluster expansion is now complete and the expanded property for a configuration $\vec{\sigma}$ can be computed very efficiently as the dot product $\vec{\omega}(\vec{\sigma})\vec{J}_{\text{final}}$.

3

Monte Carlo simulations

3.1 Monte Carlo integration

MC methods represent a broad class of computer algorithms that are based on the use of random numbers to sample high-dimensional functions. MC integration is one example of such a technique where the average value of an integral

$$I = \int_0^1 dx f(x) = \langle f(x) \rangle \tag{3.1}$$

can be approximated by evaluating $f(x)$ at N points x_i chosen at random with uniform probability over the interval $[0, 1]$. The mean value becomes

$$I_N = \langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} f(x_i) = \frac{1}{N} \sum_{i=1}^{N} f_i, \tag{3.2}$$

and the variance

$$\sigma_f^2 = \langle f^2 \rangle - \langle f \rangle^2 . \tag{3.3}$$

Such that the integral is approximated by

$$I = I_N \pm \frac{\sigma_f}{\sqrt{N}}. \tag{3.4}$$

By increasing the number of points N the error of the approximation becomes smaller and as $N \to \infty$ one approaches the correct value of I.

3.2 Importance sampling

An alternative approach to decreasing the error is to choose the points x_i based on the magnitude of $f(x)$. Consider a probability density function $p(x) > 0$ that is positive and normalized to 1 on $[0, 1]$. We can rewrite Eq. (3.1) to obtain

$$I = \int_0^1 dx f(x) = \int_0^1 dx \frac{f(x)}{p(x)} p(x) = \int_0^1 dx g(x) = \langle g(x) \rangle_p , \tag{3.5}$$

where $g(x) = f(x)/p(x)$ and the notation $\langle ... \rangle_p$ signifies that the average is obtained from sampling values of x according to $p(x)$. The mean value of the integral can then be written as

$$I_N = \langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} g_i \tag{3.6}$$

and I can be approximated with

$$I = I_N \pm \frac{\sigma_g}{\sqrt{N}}. \tag{3.7}$$

Assume now that $p(x)$ follows the approximate behavior of $f(x)$, i.e. $p(x)$ has high probability density when $|f(x)|$ is large and vice versa. Then $g(x)$ will become a smoother function than $f(x)$ and $\sigma_g < \sigma_f$. This approach of choosing values of x where $|f(x)|$ is large is called importance sampling.

3.3 Thermodynamic averages on a lattice

Now, instead of approximating a one-dimensional integral let us consider a binary lattice $\vec{\sigma}$ at a certain temperature T and find the average of some quantity A that depends on the configuration. Let the atomic lattice vector $\vec{\sigma}$ have dimensionality N, where N is the number of lattice points and the elements of the vector can assume values of either 0 or 1. The average value is then

$$\langle A \rangle = \frac{1}{M^N} \sum_{\text{all possible } \vec{\sigma}} A(\vec{\sigma}) P(\vec{\sigma}). \tag{3.8}$$

The probability of finding the system at $\vec{\sigma}$ is $P(\vec{\sigma})$, which is defined as

$$P(\vec{\sigma}) = \frac{\exp\left(-U(\vec{\sigma})/k_B T\right)}{\sum_{\text{all possible } \vec{\sigma'}} \exp\left(-U(\vec{\sigma'})/k_B T\right)}, \tag{3.9}$$

where $U(\vec{\sigma})$ is the internal energy of the configuration and k_B is the Boltzmann constant. Unfortunately it is impossible to carry out the summation in Eq. (3.8)

exactly. For illustration, consider that for a system with 100 sites the number of possible configurations, ignoring symmetry, is $2^{100} \approx 10^{30}$. Also assume that each evaluation of $A(\vec{\sigma})$ requires only one floating point operation (FLOP). The time to calculate the sum by using the largest computer cluster available [1] would still require around 10^6 times longer than the age of the universe. This demonstrates the necessity to find efficient methods to solve to Eq. (3.8).

3.4 The Metropolis algorithm

A naive approach to approximating Eq. (3.8) via a MC approach is to generate a large number of different configurations, $\vec{\sigma}$, where each element is randomly assigned a value of either 0 or 1, and estimating the average. This approach would, however, converge very slowly as the probability function $P(\vec{\sigma})$ is commonly a very sharp function, i.e. $P(\vec{\sigma})$ is close to zero for most choices of $\vec{\sigma}$. A more prudent approach would be to try to apply the importance sampling approach introduced earlier. A direct implementation of importance sampling is difficult, however, since it is not obvious how to efficiently generate configurations $\vec{\sigma}$ according to a suitable probability. Instead, an efficient way to generate configurations is to implement a so-called Markov chain, where each new configuration generated is based on a probability ratio that depends on the previous configuration.

The first adaptation of such a approach was described in 1953 by Metropolis *et al.* to determine the equation of state for a hard sphere liquid [20]. It is based on the understanding that thermodynamic averaging only requires knowledge of relative rather than absolute probabilities such as in Eq. (3.8). It uses a Markov chain to generate configurations that are more important by rejecting configurations that are unlikely, similar to importance sampling.

3.4.1 Markov chain Monte Carlo

A Markov chain is a stochastic process, in which a system undergoes transitions from one state to another. The Markov process is characterized by a lack of memory of where it has been. The future of the chain depends solely on the current state. The transition probability is given by a transition matrix

$$T(X \to X') \equiv T_{XX'}, \tag{3.10}$$

for a transition from state X to state X'. The transition probability has to satisfy

$$0 \leq T_{XX'} \leq 1 \tag{3.11}$$

[1] The IBM Summit systems at Oak Ridge National Laboratory is currently the largest non-distributed computer cluster with a peak performance of about 200 PFLOPS.

and the probability has to be normalized

$$\sum_{X'} T_{XX'} = 1. \tag{3.12}$$

The task is to generate a Markov chain of configurations such that they have a distribution proportional to the Boltzmann factor and this distribution should be independent of the position in the chain and the initial configuration. The Markov chain can exhibit these properties under certain conditions, at least for a sufficiently long time, so that the configuration can loose memory of its initial state. These conditions are:

- The Markov chain needs to be irreducible, i.e. every configuration included in the ensemble should be accessible from every other configuration within a finite number of steps.
- There should be no periodicity. Periodicity means that it is not possible to revisit a configuration except after $t = nk$ steps, $n = 1, 2, 3 \ldots$, where k is fixed.

A Markov chain that satisfies these conditions is called ergodic. If the Markov chain is ergodic it converges to a unique stationary distribution. The transition probability needs to be chosen such that the stationary distribution is the desired distribution. To assure this, consider the stationary distribution $\rho(X)$; one can also introduce a new function $\rho(X, t)$, which gives the probability of finding configuration X after t Markov steps, which for an ergodic chain becomes independent of t if t is large. This function can change from one step to another by

- going from X at step t to X' at $t + 1$ leading to a decrease in $\rho(X)$
- going from X' at step t to X at $t + 1$ leading to an increase in $\rho(X)$

that can be summarized with

$$\rho(X, t+1) - \rho(X, t) = -\sum_{X'} T(X \to X')\rho(X, t) + \sum_{X'} T(X' \to X)\rho(X', t). \tag{3.13}$$

This equation is called the master equation. The stationary solution of this equation is found by requiring $\rho(X, t + 1) = \rho(X, t)$ so we have

$$\sum_{X'} T(X \to X')\rho(X, t) = \sum_{X'} T(X' \to X)\rho(X', t). \tag{3.14}$$

Leaving out the t-dependence, which is permissible thanks to the "memory loss" of a Markov chain, yields

$$T(X \to X')\rho(X) = T(X' \to X)\rho(X'), \tag{3.15}$$

which is known as the condition of detailed balance. This means that in equilibrium the average number of steps that results in the system leaving state X must be exactly equal to the number of steps from all other states X' to X. This means that $\rho(X)$ and $\rho(X')$ do not change. Since this is true for all pairs of X and X' the probability distributions will remain stationary. Reformulating the detailed balance condition with the transition probability in this form yields

$$T(X \to X') = \omega_{XX'} A_{XX'}, \tag{3.16}$$

where $\omega_{XX'}$ is the probability for going from state X to state X' and is symmetric $\omega_{XX'} = \omega_{X'X}$. $A_{XX'}$, which must lie between 0 and 1, is the acceptance probability for actually performing the change. The detailed balance condition can then be expressed as

$$\frac{A_{XX'}}{A_{X'X}} = \frac{\rho(X')}{\rho(X)}. \tag{3.17}$$

If the sought after distribution is the Boltzmann distribution, $\rho(X) = \exp -\beta U(X)$, there are a number of choices for the acceptance probability $A_{XX'}$ that will accomplish this. The choice of Metropolis *et al.* was

$$\begin{cases} A_{XX'} & = \exp\{\beta\left[U(X') - U(X)\right]\} \quad \text{if } \rho(X') < \rho(X) \\ A_{XX'} & = 1 \qquad\qquad\qquad\qquad\quad \text{if } \rho(X') \geq \rho(X). \end{cases} \tag{3.18}$$

The Metropolis algorithm can now be formulated as follows:

- Starting from a state X, make a small trial move into a new state X' with a probability of $\omega_{XX'}$.
- Compare the weights of the distribution for the different states $\rho(X)$ and $\rho(X')$. $A_{XX'}$, the acceptance probability, is chosen equal to 1 if $\rho(X') > \rho(X)$ else it is chosen to be equal to $\rho(X')/\rho(X)$.
- The new state X' is accepted with probability $A_{XX'}$ (the system moves from X to X') and is rejected with probability $1 - A_{XX'}$ (the system remains in state X). To decide if a state is accepted or not, a random number is generated uniformly in the range $[0, 1)$ and compared to the acceptance probability. If the random number is larger than the acceptance probability the trial move is accepted.

Since each trial move is only a small change in the configuration there is an inherent correlation between the states X and X'. There is thus a correlation length s for the Markov chain and it is necessary to carry out s trial steps before reaching a new uncorrelated configuration. One MC sweep (or cycle) is defined as N trial steps, where N is the number of particles, which is commonly used to approximate s. There is also a need to equilibrate the initial configuration meaning that it is necessary to run the Metropolis algorithm before the actual sampling commences.

3.5 Probability of a state

The sampling procedure described in the previous section requires the probability ratio of two microstates. Let us consider again the rigid binary lattice introduced above. Each possible combination of zeros and ones on the $\vec{\sigma}$ vector describes a microstate of the system. Consider the case where this system is in thermal contact with a heat reservoir with constant temperature and an infinitely large heat capacity, i.e. we can add and remove as much energy as necessary without affecting the temperature of the reservoir. Since the composition of the system might change there can also be an exchange of atoms between system and reservoir.

For an isolated system, all possible microstates are equally probable. The system of N sites considered here is, however, not isolated since it is in contact with a heat reservoir. Rather the *joint* system and the heat reservoir form an isolated system. The joint microstates of system and heat reservoir will therefore be equally probable.

Consider two microstates s_1 and s_2 with a corresponding number of accessible states $\Omega_R(s_1)$ and $\Omega_R(s_2)$ for each respective reservoir. At this point there is no way of telling what the actual probabilities of the different states are since the number of accessible states associated with the heat reservoirs is unknown. All we know at this point is that the probability of a state is proportional to the total number of accessible microstates, i.e. $P(s_1) \propto \Omega_R(s_1)$. Consider now the ratio of probabilities

$$\frac{P(s_1)}{P(s_2)} = \frac{\Omega_R(s_1)}{\Omega_R(s_2)}. \tag{3.19}$$

Rewriting this equation by using the definition of entropy $S = k_B \ln \Omega$ one obtains

$$\frac{P(s_1)}{P(s_2)} = \frac{e^{S_R(s_1)/k_B}}{e^{S_R(s_2)/k_B}} = e^{[S_R(s_1) - S_R(s_2)]/k_B}. \tag{3.20}$$

Now the ratio of probabilities depends on the change of entropy in the reservoir when going from state s_1 to state s_2. The change in the entropy of the reservoirs should be small since the system is small compared to the reservoir. Then we can use the thermodynamic identity

$$dS = \frac{1}{T}\left(\mathrm{d}U + P\mathrm{d}V - \mu \mathrm{d}N\right). \tag{3.21}$$

Since all ensembles discussed in this thesis have constant volume V and fixed number of total atoms (or sites) N, the expression can be simplified to $dS = S(s_2) - S(s_1) = \frac{1}{T}\left(\mathrm{d}U - \Delta\mu\Delta N_A\right) = -\left(E(s_2) - E(s_1) - \Delta\mu(N_A(s_2) - N_A(s_1))\right)$, where E is the internal energy of the system, $N_A(s_2)$ and $N_A(s_1)$ are the numbers of A

atoms in system s_2 and s_1, respectively, and $\Delta\mu$ is the chemical potential difference between species A and B

$$\frac{P(s_1)}{P(s_2)} = e^{(E(s_2)-E(s_1)-\Delta\mu\Delta N_A)/k_BT}. \tag{3.22}$$

As was shown in section section 3.5 this ratio of probabilities is all that is needed to carry out a Metropolis MC simulation.

For the sake of completeness, however, let us determine the actual probability of the microstate. First we separate all terms in Eq. (3.22) related to s_1 to one side and terms related to s_2 to the other side,

$$P(s_1)e^{(E(s_1)-\Delta\mu N_A(s_1))/k_BT} = P(s_2)e^{(E(s_2)-\Delta\mu N_A(s_2))/k_BT}. \tag{3.23}$$

The right hand side does not depend on state s_1 and the left hand side is indepdent of s_2 and must therefore be equal to a constant

$$P(s_1)e^{(E(s_1)-\Delta\mu N_A(s_1))/k_BT} = A \tag{3.24}$$

and the probability becomes

$$P(s_1) = Ae^{-(E(s_1)+\Delta\mu N_A(s_1))/k_BT}. \tag{3.25}$$

Furthermore, the probabilities of all states must sum up to 1

$$\sum_{s_i} P(s_i) = A\sum_{s_i} e^{-(E(s_i)+\Delta\mu N_A(s_1))/k_BT} = 1, \tag{3.26}$$

which leads to

$$A = \frac{1}{\sum_{s_i} e^{-(E(s_i)+\Delta\mu N_A(s_i)))/k_BT}}. \tag{3.27}$$

Thus we see that the normalization constant A is equal to $1/\mathcal{Z}$ where $\mathcal{Z}$ is the partition function and the probability of state s_1 becomes

$$P(s_1) = \frac{e^{-(E(s_1)+\Delta\mu N_A(s_1))/k_BT}}{\mathcal{Z}} = \frac{e^{-(E(s_1)+\Delta\mu N_A(s_1))/k_BT}}{\sum_{s_i} e^{-(E(s_i)+\Delta\mu N_A(s_i)))/k_BT}}. \tag{3.28}$$

3.6 The canonical ensemble

In the canonical ensemble the volume V, the temperature, T, and the number of atoms for each species N_i are fixed. Equation (3.28) is then simplified so that the probability of a state only depends on its energy,

$$P_{\text{canonical}}(s_1) = \frac{e^{-E(s_1)/k_BT}}{\sum_{s_i} e^{-E(s_i)/k_BT}}. \tag{3.29}$$

The probability ratio used in the MC simulation becomes

$$\frac{P_{\text{canonical}}(s_1)}{P_{\text{canonical}}(s_2)} = e^{(E(s_2)-E(s_1))/k_B T}. \tag{3.30}$$

Since the number of atoms of each kind is kept fixed the only configurations being explored in the MC simulations correspond to a re-ordering of atoms from the initial configuration. Therefore in lattice based models trial steps consist of swapping the species between two sites.

3.7 Semi-grand canonical ensemble

In the semi-grand canonical (SGC) ensemble the volume V, the temperature T, the chemical potential difference(s) $\Delta\mu_i$, and the number of sites N are fixed but the relative composition is allowed to change. Equation (3.28) is then kept as is and the probability of a state becomes (for a binary system)

$$P_{\text{SGC}}(s_1) = \frac{e^{-(E(s_1)+\Delta\mu N_A(s_1))/k_B T}}{\sum_{s_i} e^{-(E(s_i)+\Delta\mu N_A(s_i)))/k_B T}}. \tag{3.31}$$

The probability ratio used in the MC simulation becomes

$$\frac{P_{\text{SGC}}(s_1)}{P_{\text{SGC}}(s_2)} = e^{(E(s_2)-E(s_1)-\Delta\mu\Delta N_A)/k_B T}, \tag{3.32}$$

In the SGC ensemble the concentrations (yet not the total number of sites) are allowed to change. Therefore the trial step consists of selecting a site and changing its occupation to another species.

Unlike the canonical ensemble the SGC ensemble allows one to access the derivative of the configurational free energy derivative via

$$\frac{\partial F}{N\partial \langle x \rangle} = \Delta\mu, \tag{3.33}$$

where $\langle x \rangle$ is the average concentration observed in the MC simulation for a certain temperature and chemical potential difference. By varying the chemical potential difference the free energy derivative for all concentrations that fall into single-phase regions (see below for the importance of this limitation) can be obtained and the configurational free energy can be found via

$$F(x,T) = E_0 + \int_0^x \Delta\mu \mathrm{d}c. \tag{3.34}$$

3.8 Variance constrained semi-grand canonical ensemble

The variance constrained semi-grand canonical (VCSGC) ensemble [21] is similar in many regards to the SGC ensemble. Both ensembles allow sampling across the entire concentration axis. In systems with a miscibility gap the SGC ensemble is, however, unable to sample inside the miscibility gap due to the mapping of chemical potential difference to concentration is multi-valued [21]. Due to the constraint in concentration the VCSGC ensemble is able to sample in miscibility gaps.

When sampling in the VCSGC ensemble the trial move consists of changing the species of a random site as in the case of the SGC ensemble. This trial move is associated with a probability,

$$P = \min\left\{1,\, \exp\left[-\beta\Delta E - \kappa\Delta N_B(\phi + \Delta N_B/N + 2N_B/N)\right]\right\}, \tag{3.35}$$

where ΔE is the energy change associated with the trial move, $\beta = (k_B T)^{-1}$, ΔN_B is the number of particles of type B, N is the total number of sites, and ϕ and κ are the average and variance constraint parameters. One of the great benefits of the VCSGC ensemble is that typically one can keep $\kappa = 200$ and vary ϕ from -2.2 to 0.2 to cover the entire concentration axis. In the SGC ensemble the $\Delta\mu$ values varies between systems and will thus require more hands-on testing to sample across the entire concentration axis.

In the VCSGC ensemble the first derivative of the free energy is associated with the thermodynamical average of concentration $\langle c_B \rangle = N_B/N$ by,

$$\beta\frac{\partial F}{\partial c_B} = \kappa\left(\phi + 2\langle c_B\rangle\right). \tag{3.36}$$

Mapping out the entire concentration axis, $\langle c_B \rangle$ as a function of ϕ for a fixed κ therefore allows one to recover the free energy also in two-phase regions.

3.9 Validity of lattice based models

The partition function that has been discussed in this chapter only considers the summation over occupations $\vec{\sigma}$. The CE technique namely only considers the occupation vector and not displacements of the atoms. For real materials, however, the atoms also undergo vibrations around their equilibrium positions. This vibrational degrees of freedom (DOFs) are thus neglected, except for relaxations, in the CE, yet they can play a role in the actual value of a thermodynamical average. This might raise concerns about the validity of the calculated thermodynamical averages using CEs to represent the Hamiltonian in MC simulations. This section

therefore considers the approximations made when using a CE in MC simulations for calculating thermodynamical averages.

In the canonical ensemble (section 3.6) the partition function is defined as

$$\mathcal{Z}(N,V,T) = \sum_{\text{all states j}} \exp\left[-\beta E_j(N,V,T)\right]. \tag{3.37}$$

The partition function involves a sum over all microstates j of the system, each of which can be split up into one part consisting of the ordering of the atoms on a lattice and one part consisting of all possible displacements of the atoms away from their equilibrium positions for that specific ordering

$$\mathcal{Z}(N,V,T) = \sum_{\{\vec{\sigma}\}} \sum_{\{\nu\in\vec{\sigma}\}} \exp\left[-\beta E\left(\vec{\sigma},\nu,N,V,T\right)\right], \tag{3.38}$$

where $\{\vec{\sigma}\}$ is the set of all possible $\vec{\sigma}$ with constant N and $\{\nu \in \vec{\sigma}\}$ is the set of all displacements of the atoms which project on $\vec{\sigma}$. Thus $\{\nu \in \vec{\sigma}\}$ represents a subspace of the phase space of the original ensemble. The partition function can then be written as

$$\mathcal{Z}(N,V,T) = \sum_{\{\vec{\sigma}\}} \Lambda(\vec{\sigma},N,V,T), \tag{3.39}$$

with

$$\Lambda(\vec{\sigma},N,V,T) = \sum_{\{\nu\in\vec{\sigma}\}} \exp\left[-\beta E\left(\vec{\sigma},\nu,N,V,T\right)\right]. \tag{3.40}$$

In other words $\Lambda(\vec{\sigma},N,V,T)$ is the partition function for the subspace of the partition function of the full ensemble for which all microstates project to the same configuration $\vec{\sigma}$. We can associate a free energy, $W(\vec{\sigma},N,V,T)$, to this partition function as

$$W(\vec{\sigma},N,V,T) = -\frac{1}{\beta}\ln\Lambda(\vec{\sigma},N,V,T), \tag{3.41}$$

which allows us to write the canonical partition function as

$$\mathcal{Z}(N,V,T) = \sum_{\{\vec{\sigma}\}} \exp\left(-\beta W(\vec{\sigma},N,V,T)\right). \tag{3.42}$$

Consider now the Hamiltonian $W(\vec{\sigma},\nu,N,V,T)$, which represents the free energy of a system, in which there are only displacive DOFs. For a given $\vec{\sigma}$ it can be expressed as

$$W(\vec{\sigma},\nu,N,V,T) = U^0(\vec{\sigma},N,V,T) + F_{\text{ex}}(\vec{\sigma},\nu,N,V,T), \tag{3.43}$$

where U^0 is the energy of the static configuration in the $\{\nu \in \vec{\sigma}\}$ subspace with the lowest energy, i.e. the fully relaxed configuration and F_{ex} will then contain the remaining part of the free energy. The probability of a state described by $\vec{\sigma}$ is then

$$P(\vec{\sigma}, N, V, T) = \frac{\exp\left(-\beta\left[U^0(\vec{\sigma}, N, V, T) + F_{\text{ex}}(\vec{\sigma}, N, V, T)\right]\right)}{\mathcal{Z}(N, V, T)}. \tag{3.44}$$

As it has been shown in this chapter, in order to use P for calculating thermodynamical averages one needs to evaluate ratios of probabilities P,

$$\frac{P(\vec{\sigma}_1, N, V, T)}{P(\vec{\sigma}_2, N, V, T)} = \frac{\exp\left(-\beta\left[U^0(\vec{\sigma}_1, N, V, T) + F_{\text{ex}}(\vec{\sigma}_1, N, V, T)\right]\right)}{\exp\left(-\beta\left[U^0(\vec{\sigma}_2, N, V, T) + F_{\text{ex}}(\vec{\sigma}_2, N, V, T)\right]\right)}. \tag{3.45}$$

$U^0\left(\vec{\sigma}, N, V, T\right)$ can now be approximated by taking the zero Kelvin value. Furthermore, in practice it is expected that F_{ex} is dominated by the vibrational free energy. A common approximation is therefore to remove the dependency of $\vec{\sigma}$ on F_{ex} and also assume that F_{ex} is a linear combination of the composition [22]

$$F_{\text{ex}}(N, V, T) = N_a F_{\text{ex},A}(T) + N_b F_{\text{ex},B}(T). \tag{3.46}$$

If F_{ex} is linear in the composition then it can be completely removed in the canonical ensemble since it cancels out in the probability ratio. For the SGC ensemble it can also for many purposes be left out completely since a linear combination of concentration is just a shift of the chemical potential by a constant. The value of $U^0(\vec{\sigma}, N, V)$ can now be calculated with a CE that maps $\vec{\sigma}$ to the fully relaxed configuration. Finally P reduces to

$$P(\vec{\sigma}, N, V, T) = \frac{1}{\mathcal{Z}} e^{-\beta U^0(\vec{\sigma}, N, V)}. \tag{3.47}$$

A thermodynamic average of a property $A(\vec{\sigma}, \nu)$ at N, V, T is then approximated by the value of A for the fully relaxed configuration. This means that even if the cluster expansion acts on the perfect lattice $\vec{\sigma}$, relaxation effects are still taken into consideration when calculating $\langle A \rangle$. Yet, any vibrational or temperature dependence of a property is largely neglected in this approach. If vibrational effects are important there are ways to deal with this by for example incorporating these effects into a MC by also including the vibrational part of the free energy in a CE [23, 24]. To calculate free energies one can pick out representative structures from the trajectory in a MC simulation and calculate the vibrational free energy and add this to the configurational free energy, as done for example in Paper V.

3.10 Monte Carlo simulations example

Here, the simulation techniques discussed in this chapter will be exemplified. To that end, a fictional binary CE was made up for a face-centered cubic (FCC) system

considering only the nearest neighbor interaction. The one ECI corresponding to the nearest-neighbor pair, was chosen to be $J = -1.0$ creating a non-mixing system and energetically unfavorable A-B bonds. The Boltzmann constant k_B was set to a value of 1.0 such that temperature and energies in this section are given in units of J.

The simulation cell comprised $10 \times 10 \times 10$ primitive unit cells, and thus 1000 atoms. Each MC simulation was run for 5,000 MC sweeps, i.e. 5 million trial steps. The system has a miscibility gap that closes somewhere below $T = 2$. In Fig. 3.1 the free energy derivative as well as the free energy are shown as a function of concentration for one temperature above the critical temperature ($T = 4$) and one temperature below ($T = 1$). At $T = 1$ the SGC ensemble is unable to sample inside the miscibility gap since one value of $\Delta\mu$ maps to several compositions inside the miscibility gap [21]. The values obtained in the SGC and VCSGC ensemble are consistent for all values where SGC is able to sample.

The free energy at $T = 1$, (Fig. 3.1b), reveals that inside the miscibility gap one can lower the free energy using a tangent construction by separating into two phases, one that is rich in A and one that is rich in B. The excess free energy value for a certain concentration corresponds to the cost of forming an interface cost between these two phases. Depending on the concentration different types of interfaces are formed as the excess free energy depends both on the (facet-dependent) interface free energy γ and the interface area A (Fig. 3.2). At 5% concentration, just at the edge of the miscibility gap, small clusters of B atoms appear in the A-rich phase (Fig. 3.2 a)). At concentration 15% the shape that lowers the interface cost is a compact precipitate (Fig. 3.2 b)). As A atoms keep being added to the system one can eventually lower the interface cost by connecting the A phase to the top and bottom of the simulation box forming a pillar (Fig. 3.2 c)). Eventually, as this pillar grows in width it can connect to the sides, forming a layer, which removes the A-B interface on the sides (Fig. 3.2 d)). At this point the excess free energy no lower varies with concentration.

The entropy can simply be calculated via $S = (U - F)/T)$ (Fig. 3.1 d)).=, which shows that above the miscibility gap at $T = 4$ the entropy coincides with the ideal mixing entropy. Inside the misciblity gap, however, the system is clearly deviating from an ideal mixture.

When employing either the SGC or VCSGC ensemble a more technical aspect is revealed by looking at $\Delta\mu$ and ϕ as a function of concentration (Fig. 3.3). In the VCSGC ensemble an even sampling of the concentration axis can be simply achieved by using an even ϕ spacing, while in the case of the SGC the spacing is generally much more inhomogeneous for an evenly spaced $\Delta\mu$. Furthermore, the maximum and minimum values of $\Delta\mu$ required to sample the full range are temperature dependent. Hence, the SGC ensemble often requires some additional steps in order to use it effectively whereas with the VCSGC ensemble one can for

most systems sample the concentration axis by varying ϕ between -2.2 and 0.2 regardless of temperature.

Finally, the acceptance ratios are calculated in the different ensembles (Fig. 3.3 c)). A too low acceptance ratio is not desirable in terms of statistical and computational efficiency. The general behavior is that the canonical ensemble yields a high acceptance ratio close to the end points of the concentration axis, with the opposite behavior for the SGC and VCSGC ensembles. The VCSGC ensemble yields the highest acceptance ratio inside the miscibility gap in agreement with [25].

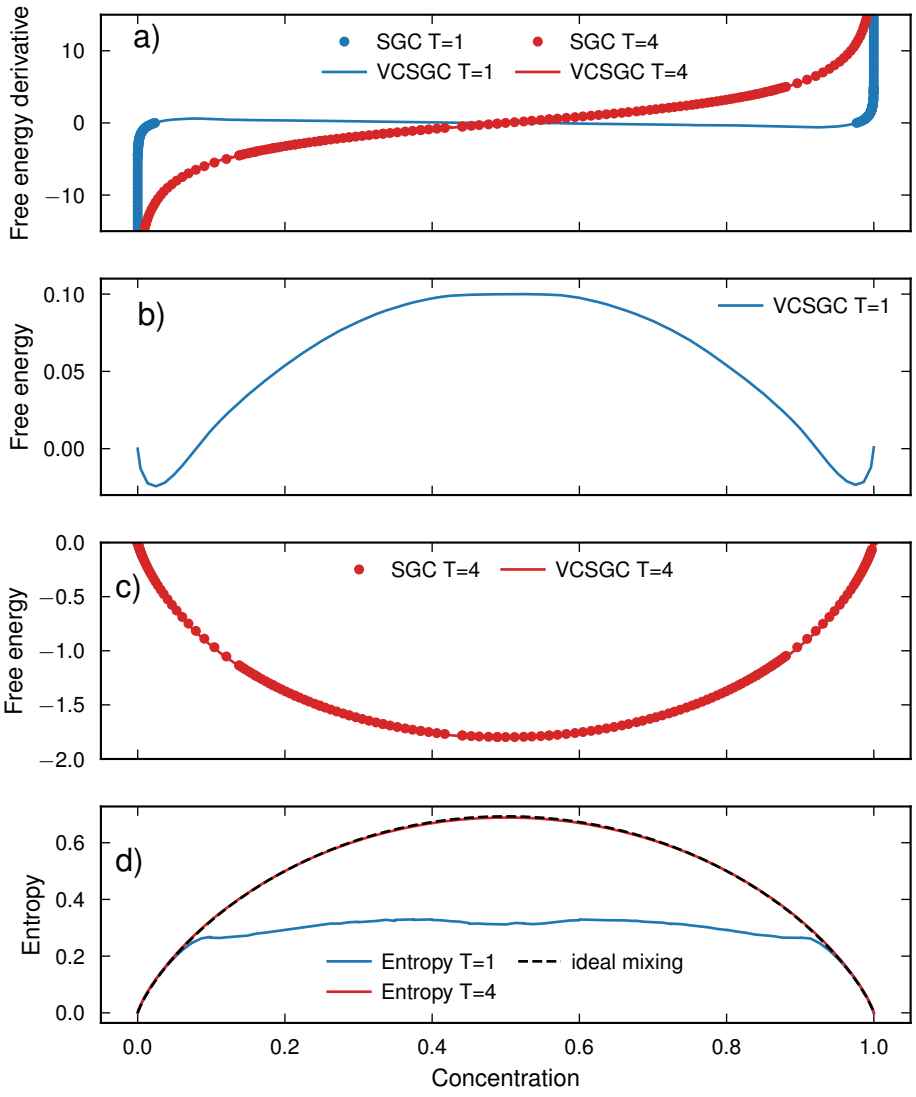

Figure 3.1: Results from MC simulations in different ensembles. In figure a) the free energy derivative as a function of concentration are shown for temperature 1 and 4. At $T = 1$ there is a miscibility gap hence the missing points for the SGC ensemble. In figure b) and c) the free energy is shown for temperature 1 and 4 respectively. At $T = 1$ the VCSGC ensemble allows constructing the free energy inside the miscibility gap. The entropy is shown in figure d). The entropy can be calculated as $S = (U - F)/T$.

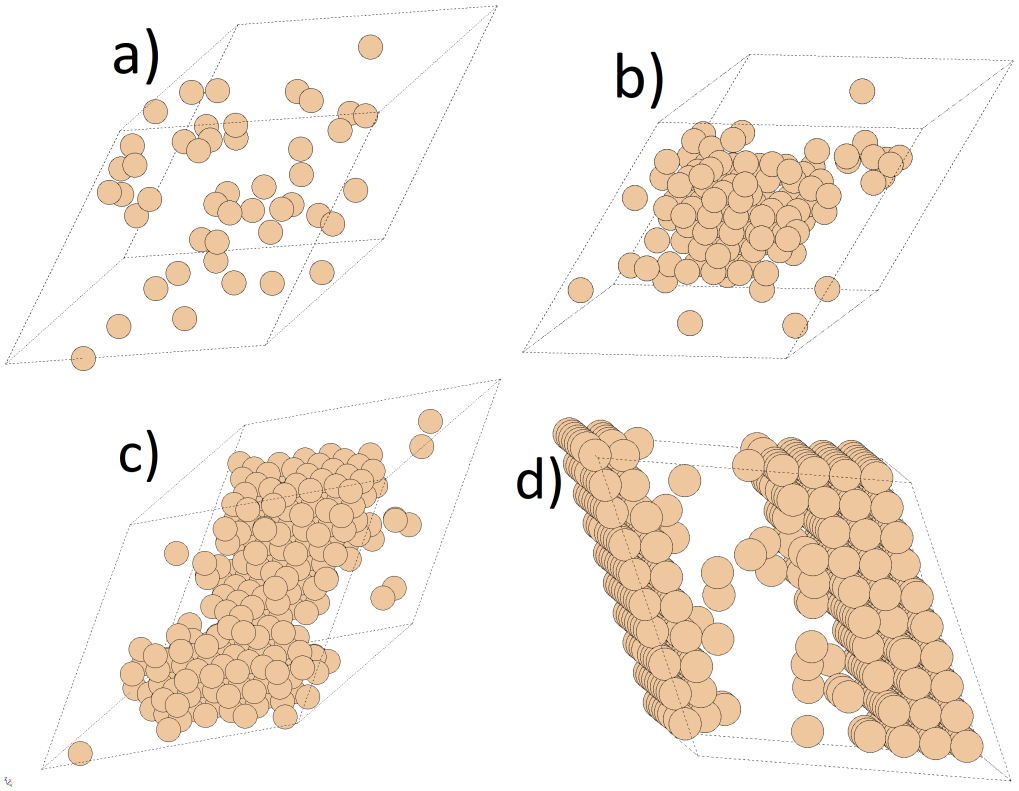

Figure 3.2: Shapes of the two-phase regions where only one species is shown. In figure a) the concentration is 5%, right at the beginning of the miscibility gap. Here the A atoms are randomly mixing in B. In figure b) the concentration is at 15% which is inside the misciblity gap, here the A rich phase is a sphere which minimizes the area to volume ratio. In figure c) the concentration is at 33 % and the A rich phase is a pillar. This minimizes the A to B interface since the top and bottom sides of the pillar do not have a B interface due to periodic boundary conditions. In figure d) the system is at 50% concentration and the shape is layered where only the face of the layer interfaces to the B phase due to periodic boundary conditions.

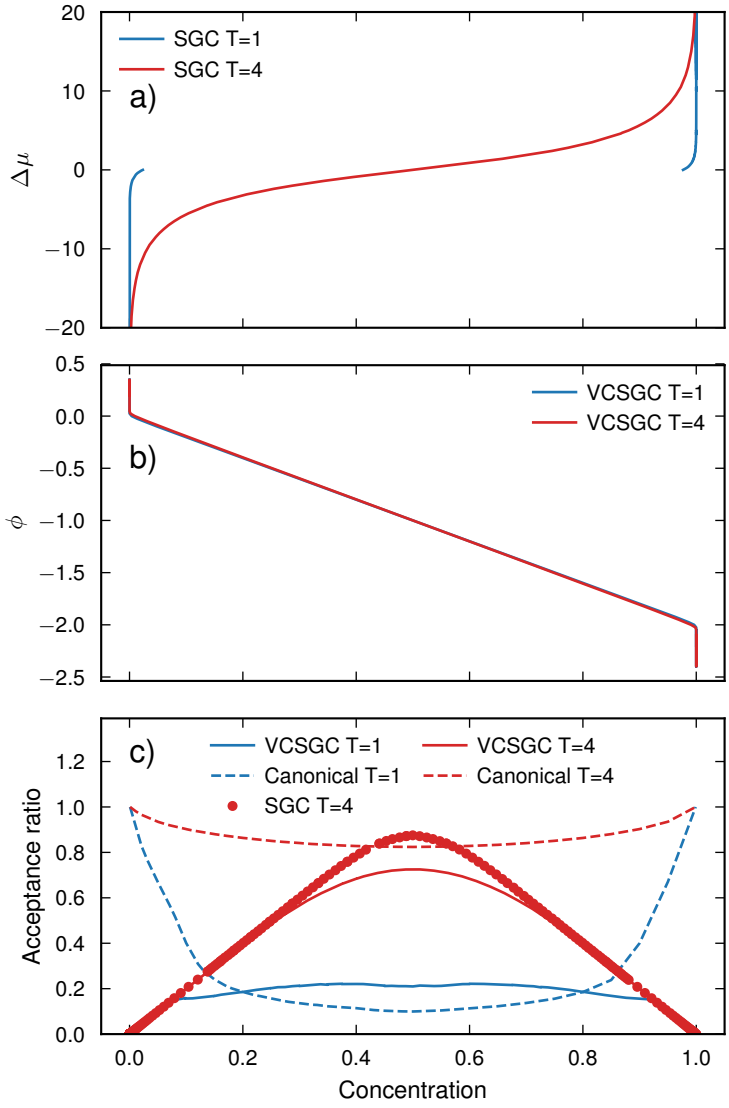

Figure 3.3: Results from MC simulations in different ensembles. In figure a) the chemical potential difference is shown as a function of concentration. The relatively steep region at low and high concentrations and flat region at intermediate concentrations makes it difficult to sample the entire concentration axis with an even spacing of $\Delta\mu$. In figure b) the ϕ parameter in the VCSGC ensemble is shown as a function of concentration. The ϕ values are linear as a function of concentration making it easy to evenly cover the concentration axis. In figure c) the acceptance ratios are shown as a function of temperature, ensemble and concentration. The general behaviour is that the canonical ensemble has high acceptance ratio at low and high concentrations and vice versa for the SGC and VCSGC ensemble.

4 Materials

4.1 Inorganic Clathrates

Inorganic clathrates constitute a class of inclusion compounds that exhibit a cage-like framework in which the cages are occupied by guest atoms or small molecules [26, 27]. The guest atoms, which are undersized relative to their respective cage, can act as so-called rattlers, lowering the lattice thermal conductivity. The framework structure can support a rather wide range of compositions, and two, three or even more components. The availability of different compositions and the resulting variability of the distribution of elements in the framework provide opportunities for optimizing material properties. The framework is typically occupied with elements from groups 13 and 14, although other combinations are also possible [27, 28]. Inorganic clathrates have been studied in particular as potential high-performance thermoelectric materials [26, 29, 30, 31]. This is due to their low intrinsic thermal conductivity [32, 29, 33], suitable band gap [34, 35, 36], good dopability, and compositional variability [37].

4.1.1 Crystal structure

Inorganic clathrate can be categorized according to their symmetry [26, 27]. The present thesis focuses on type I clathrates, which have received the most attention so far (Fig. 4.1). The framework structure of type I clathrates contains 46 tetrahedrally coordinated host atoms in the unit cell. It is the geometrical arrangement of these 46 atoms that provides eight voids (or cages) per unit cell for the guest atoms. There are two smaller dodecahedral cages and six larger tetrakaidecahedral cages. The crystal structure belongs to the cubic space group $Pm\bar{3}n$. In Wyckoff notation the center of the cages are $2a$ and $6d$ for the dodecahedral and tetrakaidecahedral

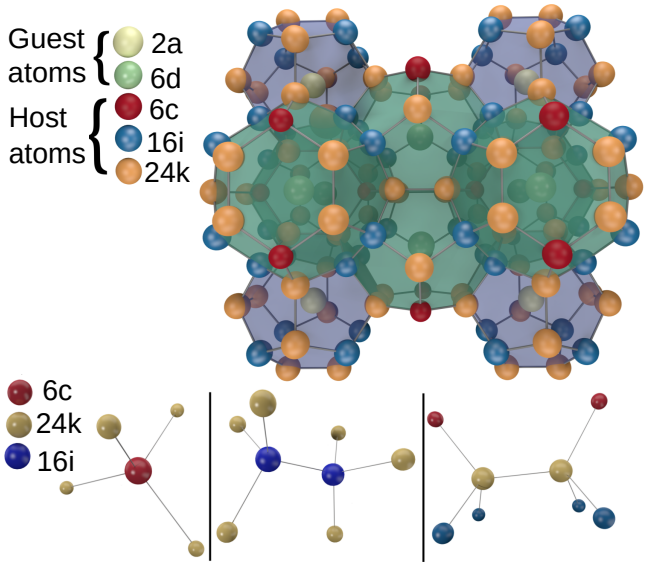

Figure 4.1: Crystal structure of type I clathrates. The guest species (Ba) occupies Wyckoff sites of type 2*a* and 6*d*, while the host species (Ga, Al, Ge, Si) occupy Wyckoff sites of type 6*c*, 16*i*, and 24*k*. The configurations in the bottom row illustrate the environments for 6*c*, 16*i*, and 24*k* sites, respectively.

cages, respectively, whereas the framework atoms occupy Wyckoff sites 6*c*, 16*i*, and 24*k*.

4.1.2 The Zintl concept

The Zintl concept provides a rationale for the stoichiometry of semiconducting clathrates [26, 27, 38, 39]. It requires four electrons to be available for each tetrahedrally bonded host atom, while the guest atom is assumed to donate its valence electrons to the host framework. The number of electrons required to form bonds between the 46 host atoms is thus 184. At stoichiometric compositions the material is thus fully charge balanced.

The general formula for type I clathrates is $A_8M_xM'_{46-x}$. The clathrates studied in this thesis are comprised of Ba, Ga/Al, Ge/Si for A, M and M', respectively. Each Ba atom can donate two electrons, whereas Ga/Al and Ge/Si provide three and four valence electrons, respectively. Therefore, for $x = 16$ all bonds saturated. Lowering the Ga/Al composition thus leads to electron deficiency and the material is expected to be n-doped. Likewise, increasing the Ge/Si concentration creates a p-doped material. As a thermoelectric generator requires both an n-doped and p-doped material to function, clathrates can in principle achieve both of these limits by variation of the composition.

4.1.3 Ordering

The host framework of inorganic clathrates usually comprises several different Wyckoff sites (in the present work $6c$, $16i$, and $24k$), which are occupied by several different species (here Ga, Ge, Si, Al). If the different sites were occupied statistically one would expect for a stoichiometric sample ($x = 16$) $16/46 \approx 35\%$ of the sites to be occupied by Al or Ga. Measurements of the so-called site occupancy factors (SOFs) reveal, however, dramatic deviations from this average [37]. The deviation from an entropically desirable random occupation (i.e., 35%) indicates that the interaction between the constituents plays a crucial role and the experimentally observed ordering (in form of the SOFs) arises from a competition of energy and entropy.

The SOFs vary between the compounds and can also show strong, non-monotonic variations with stoichiometry [37, 40, 41]. Furthermore, the SOFs show an impact on transport properties [42]. Hence, understanding the ordering of these materials is crucial for understanding their thermoelectric performance.

4.1.3.1 Empirical rules for SOFs

A set of guidelines for the SOFs has been formulated on the basis of a large set of experimental data [37]. They are mostly based on the environment of each Wyckoff site and the observation that direct bonds between trivalent atoms are unfavorable. There are three different bonding environments, one for each Wyckoff site. The $6c$ sites have four $24k$ sites as nearest neighbors; the $24k$ sites have one $6c$ site, two $16i$ sites and one $24k$ site as nearest neighbors; finally, the $16i$ sites have three $24k$ site and one $16i$ site as nearest neighbors.

The geometry of the lattice thus leads to the following set of rules.

1. A $6c$ site has no other $6c$ sites in its surrounding and hence $6c < 100\%$.

2. The same argument can be made with regard to the other two sites giving $24k < 50\%$ and $16i < 50\%$ since there is one $24k - 24k$ and one $16i - 16i$ pair per $24k$ site and $16i$, respectively.

3. Furthermore, $6c + 24k < 100\%$ since the $6c$ site binds to four $24k$ sites and the sum of the SOFs should be below 100%.

4. Also $16i + 24k \leq 50\%$. As pointed out in Ref. [40], however, this rule is too restricted and the condition to avoid trivalent nearest neighbors is actually $16i + 24k \leq 83.3\%$.

The violation of the last rule can also be seen in papers I and II in this thesis, where we present ground state SOFs with no trivalent nearest neighbor and $16i + 24k =$

56.25%. With that small modification, the rules, based on simple assumptions, are in good agreement with the experimental data.

4.1.3.2 Guideline for variations of SOFs

The rules described above cannot provide a direct rationale for explaining the *variations* in the SOFs as a function of composition. Different components in the framework structure will show different variations due to composition [41]. Here, atomic scale simulation, e.g., based on CEs parametrized using first-principles calculations, provide a way to obtain very detailed information [42, 43, 41]. In Ref. [41] it was found that Al-based clathrates ($Ba_8Al_xGe_{46-x}$, $Ba_8Al_xSi_{46-x}$) exhibit a strong, non-monotonic variation of the SOFs with Al content, whereas Ga-based clathrates ($Ba_8Ga_xGe_{46-x}$, $Ba_8Ga_xSi_{46-x}$) show a monotonic dependence on Ga content. It was argued that the differences observed were largely due to the Al–Al repulsion being twice as strong as the Ga–Ga repulsion. In other words, the Al-based clathrates demonstrate a more extreme SOF behavior in order to avoid costly Al–Al bonds. The observations and conclusions from paper III [41] have not been transferred to a general SOF model of clathrates. The results, however, demonstrate that information from first-principles calculations provides an efficient and accurate means to predict and rationalize ordering in these materials.

4.2 Metallic alloys and W-Ti

Metallic alloys are of great importance in technology and are continuously being developed and are of interest in basic research. Alloying often yield different properties compared to the original components. For example alloying iron with carbon creates steel, an alloy that is stronger and harder than pure iron.

A binary system can be categorized in two classes, either the bonds between the two constituents of the alloy are repulsive or attractive, corresponding to positive or negative mixing energies, respectively. Cu–Ag [44, 45, 46] and Fe–Cu [44, 47, 48, 49] are two non-mixing binary systems. In these systems one commonly observes a wide two-phase region commonly called a miscibility gap. Attractive interactions between the two constituents on the other hand yield solid solutions, as for example in Au–Ag [44, 50] and W–V [44, 51] or the formation of intermetallic phases as in Fe–Pt [44, 52, 53, 54] or Ni–Al [44, 55, 56]. Often mixing or non-mixing are symmetric in the sense that if A mixes in B then B also mixes in A. There are some notable examples where this does not happen. One example is Fe–Cr where mixing Cr in Fe is energetically favorable whereas Fe in Cr is energetically unfavorable [57, 58, 59]. This behavior can be explained in terms of the magnetic properties of Fe and Cr. Fe and Cr prefers ferro and antiferromagnetic ordering respectively.

Non-symmetric mixing behaviour can also be observed in non-magnetic systems due to asymmetry in lattice structures and their energetics. One such system is W–Ti, which is studied in this thesis.

4.2.1 W-Ti

At low temperatures pure titanium exhibits a hexagonal close packed (HCP) phase (α-Ti), which transitions into a BCC phase (β-Ti) at higher temperature, the latter of whic is dynamically stabilized [60]. Tungsten on the other hand maintains a BCC structure up to the melting point. W-HCP and other close packed phases are energetically very unfavorable and not stable during normal conditions [61, 62]. Hence in the W–Ti phase diagram there is a competition between BCC and HCP phases. Experimentally the W–Ti phase diagram is difficult to measure due to the high melting points of both Ti and W, which is coupled to slow kinetics, in particular on the W-rich side. Since experimental data points for tungsten concentrations around 30% and above are therefore only available down to 1473 K, the experimental phase diagram for W–Ti has been assessed using rather severe assumptions [63, 64, 65, 66]. Here, atomic scale simulations enable a theoretical assessment of the phase diagram, which is not limited by slow kinetics.

4.3 Tungsten Carbide

Tungsten carbide (WC) belongs to the transition metal carbides, which are characterized by high hardness and high melting points [67, 68, 69, 70]. For these reasons WC is a widely used material for hard and wear-resistant tools used e.g., for metal cutting and rock drilling [71].

The phase diagram of tungsten carbide reveals many different phases [72] (Fig. 4.2). At stoichiometric composition, WC has a hexagonal crystal structure, referred to as the δ-WC phase [73]. The cubic structure, referred to as γ-WC only becomes stable at very high temperatures and can contain up to 50% vacancies on the carbon lattice. It has been observed, however, that γ-WC can exist in the form of thin interfacial structures or as particles, at lower temperatures due to its favorable interfacial properties [74, 75, 76, 77, 78]. In cemented carbides, e.g. WC-Co, γ-WC can exist in the form of thin layers at the interface between a δ-WC particle and the Co matrix and can thus act as a grain growth inhibitor, leading to improved mechanical properties [74, 73, 75]. Hence it is of interest to understand and model the δ-WC and γ-WC phases. In this thesis the phase diagram of δ-WC and γ-WC has been constructed using computational methods.

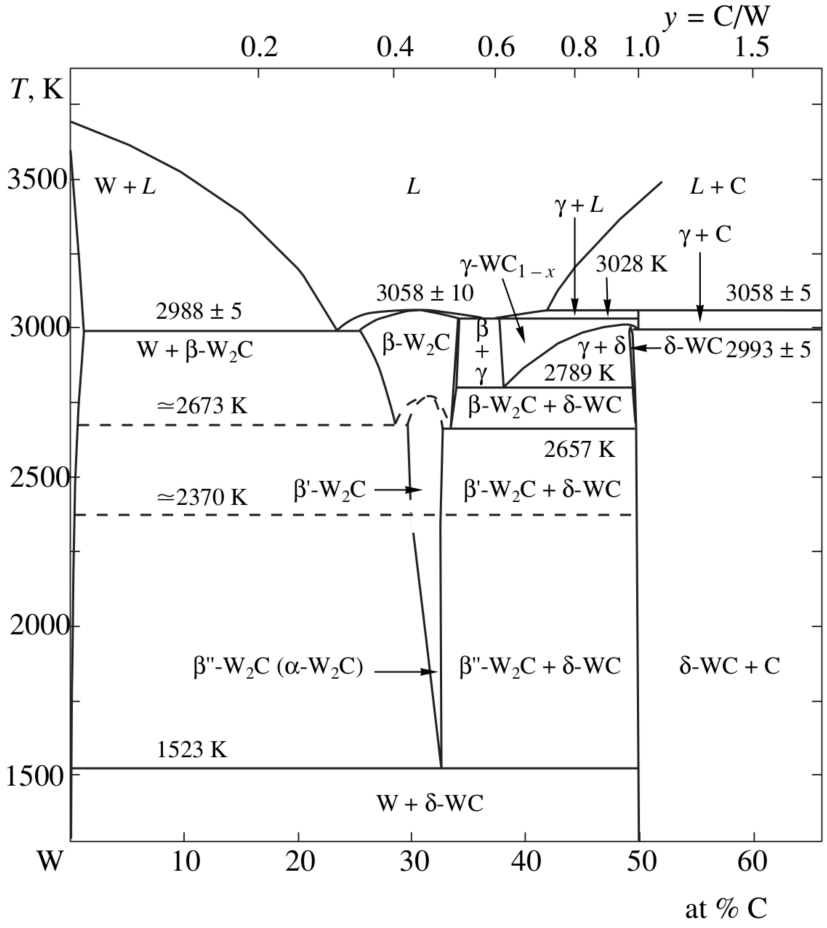

Figure 4.2: Phase diagram of the W–C system [72].

4.4 Zeolites

Zeolites are of great industrial importance due to their use as catalysts. Zeolites are aluminosilicates that exhibit a microporous crystalline framework with a wide range of pore sizes and cages (Figure 4.3). This framework is built from SiO_4W tetrahedra, at the center of which resides either an Al or Si atom surrounded by four O atoms. Each O atom is connected to two Si/Al sites and from these elemental building blocks many different zeolite frameworks can be built. Currently there more than 200 different zeolite frameworks have been recognized by the International Zeolite Association [79].

Starting from a pure framework of Si and O, a substitution of a tetravalent Si^{4+} with a trivalent Al^{3+} creates a charge imbalance. This charge imbalance is compensated by adding counterions. There can be many different counterions in

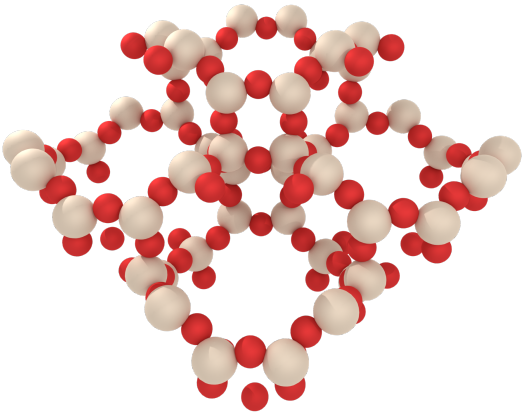

Figure 4.3: Crystal structure of the SSZ-13 chabasite. Oxygen is here depicted in red and Si/Al sites in beige.

zeolites which each yield different properties. The choice of counterion thus allows one to engineer zeolites for specific purposes. For instance, by charge compensating with H^+ one creates Brønstedt acid sites, whereas Cu^+ counterions will enable redox chemistry.

The catalytic performance of zeolites depends on the distribution of the Al atoms in the framework [80, 81] hence understanding and controlling the ordering of Al atoms is a crucial part of creating and understanding zeolites and their properties. Unfortunately, as Si is substituted with Al the number of ways the Al atoms and the counterions can be distributed grows quickly with Al content. Modeling and describing the ordering in zeolites can thus be a difficult problem.

To rationalize the Al distribution in zeolites it is common practice to refer to Löwenstein's rule [82, 83], which states that Al–O–Al motifs are forbidden, i.e. two Al are not allowed to share the same oxygen. Exceptions to this rule have, however, been noted [84]. In paper VI the Löwenstein rule is therefore critically examined.

5

Summary of the papers

5.1 Paper I

In paper I we introduced ICET, a tool for constructing and sampling CEs. To demonstrate the capabilities of ICET and CEs we applied it to two different materials. First, we demonstrated the construction of a CE for the binary alloy Ag–Pd. The CE was subsequently sampled using MC simulations in order to obtain the configurational free energy as a function of composition and temperature. Finally, the free energy was used to construct the phase diagram of Ag–Pd.

Furthermore, we considered ordering in an inorganic clathrate. An ensemble of CEs was constructed using the shuffle-split method. Each CE in the ensemble was sampled in MC simulations sampled and the SOFs were obtained by averaging over these predictions. This approach enabled us quantify the sensitivity of the predicted properties to uncertainties in the model.

5.2 Paper II

Paper II focused on the inorganic clathrate $Ba_8Ga_{16}Ge_{30}$. A CE was constructed to represent the energies of fully relaxed structures obtained from DFT calculations and subsequently sampled by MC simulations, from which the chemical order as a function of temperature was obtained, specifically the SOFs. Representative configurations for specific temperatures were extracted from the simulations and further analyzed with respect to their electrical transport properties using DFT and Boltzmann transport theory (BTT) calculations. This combination of CE and MC simulations as well as DFT and BTT allowed us to obtain the thermoelectric power factor as a function of temperature (Figure 5.1). The approach was validated

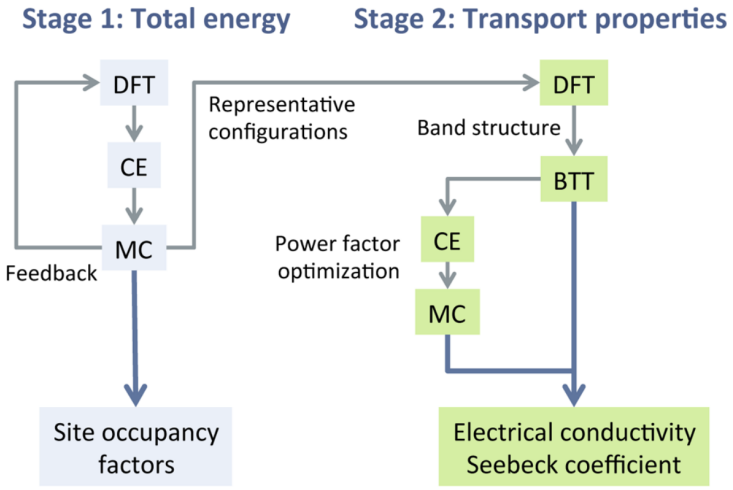

Figure 5.1: Schematic illustration of the methods employed in paper I. DFT: density functional theory; MC: Monte Carlo simulations; CE: alloy cluster expansion; BTT: Boltzmann transport theory.

by comparing the SOFs and transport coefficients with experimental data. Another CE was then constructed to represent the power factor as a function of the chemical configuration and was employed to determine the chemical ordering that maximized the power factor. The optimized structure yielded a power factor increase by more than 60%. This was achieved by reducing the number of trivalent species on the 6*c* Wyckoff site. Hence, the approach developed in this paper demonstrates the use of CEs for structure optimization.

5.3 Paper III

Paper III addressed the chemical ordering in the clathrate systems $Ba_8Ga_xGe_{46-x}$, $Ba_8Ga_xSi_{46-x}$, $Ba_8Al_xGe_{46-x}$, and $Ba_8Al_xSi_{46-x}$ as a function of composition (Figure 5.2). The ordering in these materials can significantly impact the material properties, see e.g., paper II. Hence, an understanding of the ordering is crucial in order to understand and optimize these materials. To that end, for each system a CE was constructed based on the energies of fully relaxed structures obtained from DFT calculations and the SOFs were obtained from MC simulations. The simulated SOFs agree very well with experimental data (Figure 5.2), which allowed us to clarify variations and trends in the experimental data. In particular Al-based clathrates were found to display strongly non-monotonic variations of the SOFs with composition (Figure 5.2c,d).

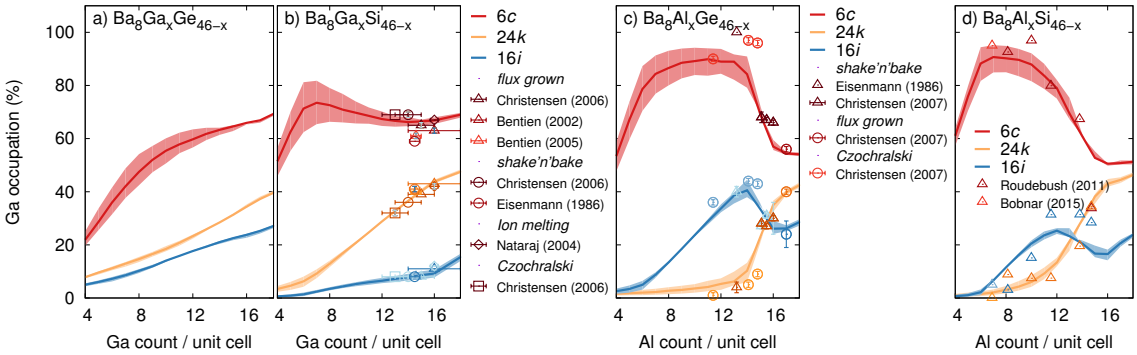

Figure 5.2: Site occupancy factors in intermetallic clathrates as a function of composition for Wyckoff sites 6*c* (red), 16*i* (blue), and 24*k* (orange). Solid lines show simulation results obtained at 700 K whereas the shaded regions indicate a variation by ± 100 K.

5.4 Paper IV

In paper IV we constructed the phase diagram of W–Ti, a system which features metastable boundary phases on both sides of the phase diagram. Titanium exhibits a transition from HCP at low temperatures to BCC at high temperatures, where the latter is stabilized by vibrations [60]. Tungsten maintains a BCC structure up to the melting point. Close-packed structures such as HCP are energetically unfavorable and only thermodynamically stable at high pressures [61, 62]. Thus in order to construct the W–Ti phase diagram we had to consider two competing crystal lattices, BCC and HCP. The computational approach was to generate sets of training structures with varying composition and ordering for both HCP and BCC. Two CEs were constructed, one for HCP and one for the BCC lattice. Each CE was sampled using MC simulations in the VCSGC ensemble in order to calculate the configurational free energy. Finally, using a model for the vibrational free energy, we were able to construct the phase diagram of W–Ti.

5.5 Paper V

In paper V, we investigated the thermodynamic properties of two WC phases, the cubic phase (γ-WC) and the hexagonal phase (δ-WC). In order to study these phases we considered configurational, vibrational, and electronic DOFs by combining effective harmonic models as implemented in the HIPHIVE[85, 10] package and alloy CEs as implemented in the ICET package (Paper I).

The hexagonal phase contains less than 1% carbon vacancies and can thus be modeled in the dilute limit. The γ-WC phase can have vacancy concentrations up

to 50% and is only thermodynamically stable close to the melting point. To model the carbon-vacancy lattice in γ-WC a CE was constructed. This CE was sampled in the SGC ensemble, which is essentially the grand canonical (GC) ensemble since we are dealing with vacancies, in order to obtain the configurational free energies as well as finding ground states and representative configurations needed for the modeling of vibrational and electronic free energy.

It is found that stoichiometric γ-WC is dynamically unstable in the harmonic approximation but stabilized when introducing up to around 25% carbon vacancies. This alone demonstrates the challenges faced when modeling the configurational and vibrational DOFs in γ-WC. In order to model the vibrational DOF effective harmonic model (EHM) were constructed from ab initio molecular dynamics simulations at various temperatures. The vibrational free energy of mixing was calculated across the entire concentration range and was found to be anti-correlated with the mixing energy. Hence the inclusion of vibrational free energy in this system is not purely an offset to the total free energy but has a strong concentration dependence and effectively flattens out the free energy of mixing. We also observed significant differences in the vibrational properties between ground state and representative (high temperature) configurations. Eventually, the phase diagram was found by tangent constructions with the combined free energies of the different phases for each temperature (Figure 5.3). Our calculated transition temperatures are 5 to 10% lower than in experimental studies but the overall phase diagram is in good agreement.

While the electronic contributions to the free energy are smaller than configurational and vibrational ones, we demonstrate that all DOFs are important for understanding the thermodynamic stability of the cubic phase. We also note that thermal expansion has a non-negligible effect on the final free energies and phase diagram. While the cubic phase is found to be stable at temperatures above 2500 K, our calculations reveal that there exist many ordered low temperature structures, e.g. W_8C_7 and W_8C_4, similar to other transition metal carbides[86], which may be of important when considering, e.g., cubic interfacial structures or nanoparticles.

5.6 Paper VI

Löwenstein's rule state that Al–O–Al motifs are energetically very unfavorable, i.e. two Al atoms should not share the same O atom. In paper VI we examined the Löwenstein's rule in zeolites and demonstrated that it works as expected when applied in its original context. We identify the conditions under which it falls short and formulate a generalized version that explain both prior observations of violations of Löwenstein's rule and how one can achieve Al–O–Al motifs in principle.

As a prototype we considered the SSZ-13 chabasite system due to its small unit

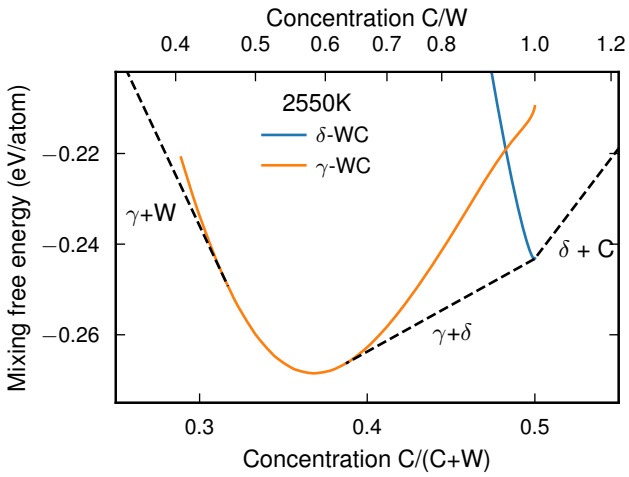

Figure 5.3: Mixing free energy for δ-WC and γ-WC phases as a function of C concentration at 2550 K. The convex hull is indicated by a dashed black line. δ indicates δ-WC, γ indicates γ-WC, W indicates W BCC, and C indicates graphite.

cell, which simplified the analysis. In zeolites as one substitutes Si^{4+} with Al^{3+}, a charge deficit is created which is compensated by adding counterions (H^+, Na^+, K^+, Rb^+). The number of ways one can distribute Al along with the counterions quickly becomes a combinatorically challenged problem as the Al content increases. This challenge makes it impractical to study these systems using enumeration and it also is an indication on how the configurational entropy might matter in these systems. We thus constructed CEs which allow accurate and efficient computation of the energy for all distributions of Al and counterions. For each counterion a CE was constructed using around 100 to 200 training structures as calculated from DFT. The CV score ranged from 1.7 meV/atom for Rb to 6.6 meV/atom for the system with free charge carrier compensation. The systems were sampled with MC simulations in the VCSGC ensemble in order to get thermodynamical averages of the fraction of nearest neighbor Al–Al pairs (i.e. Al–O–Al motifs) as a function of composition.

The results reveal very different behavior depending on the counterion species (Figure 5.4). With counterions Na^+, K^+, and Rb^+ there are almost no Al–Al nearest neighbors up to about 33% Al. The behavior of these counterions is thus largely consisten with Löwenstein's except for some small amount of Al-Al bonds at lower Al concentration, which are stabilized by entropy.

With H^+ as counterion, however, the picture is reverse (Figure 5.4). Here, the fraction of Al–Al bonds increases rapidly with increasing Al content to 40% Al-Al nearest neighbor bonds and from there continues to increase gradually. The Al

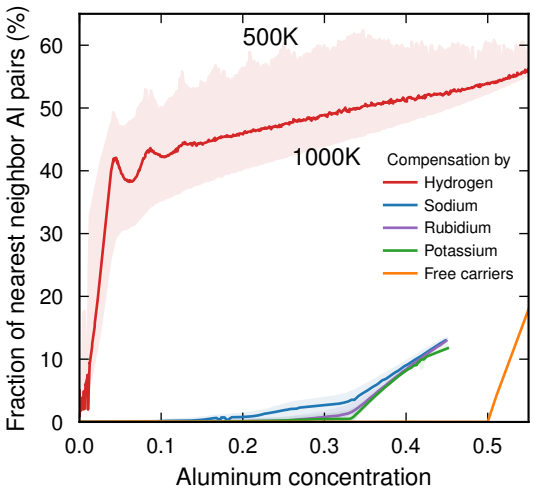

Figure 5.4: Variation of the fraction of Al–Al nearest-neighbor pairs as a function of Al content using either ions (H^+, Na^+, K^+, Rb^+) or free carriers for charge compensation.

atoms are thus effectively attracted to other Al atoms when being charge compensated by hydrogen, which is also noted in Ref. [84].

In paper VI we rationalize this contrast between H^+ and the other counterions by considering counterion size, level of charge localization, and local structure rearrangements.

Finally, we note that the approach taken in paper VI can be extended to other zeolite structures as well as other counterions. The rich choice of different frameworks and counterions in zeolites offers both a challenge in terms of modeling but also an opportunity in finding optimal zeolites.

6

Conclusions and outlook

6.1 Conclusions

This thesis was concerned with the study of ordering phenomena in several different materials. The method of choice has been CEs, which has proven to be a very efficient and accurate approach.

When applied to inorganic clathrates (papers II and III) or zeolites (paper VI) these models achieved very good agreement with available experimental results. More interestingly the microscopic simulations provided novel insight in the ordering in these materials that is inaccessible by experiments alone.

In the latter cases, the underlying lattice was vibrationally stable in the relevant concentration range, whence CEs alone were sufficient to capture the leading contributions to the free energy. This is not always the case as there are many material systems, in which at least one of the relevant phases is only stabilized by vibrations. In this thesis, CEs were therefore combined with dynamic models to account for the configurational and vibrational contributions to the free energy, respectively. In the case of the W–Ti system (paper IV) CEs were combined with data from *ab-initio* molecular dynamics (MD) simulations and thermodynamic data from CALPHAD to obtain the full phase diagram. In the case of W–C (paper V) this approach was taken further as cluster and force constant expansions were combined to analyze the counter effects of configurational and vibrational DOFs.

There are, however, some caveats with this method. The Hamiltonian used in the MC simulations should in principle also consider the vibrational free energy, such that the configurations visited in the MC trajectory depend on the vibrational free energy. The methodology used in this thesis could thus be further improved by allowing a simultaneous sampling of the vibrational and configurational DOFs.

One way could be to construct a CE for the vibrational free energy [23], another could be to create a combined cluster and force constants expansion [87].

6.2 Outlook

Looking forward there are many ways, one may build on the results of this thesis. Regarding the clathrates there is an abundance of different species one can consider on both the framework but also as guest atoms. The search for novel thermoelectric materials could thus start from paper II where we constructed a CE for the power factor and found the configuration with optimal properties. In principle, one can repeat the calculations done in paper II in a high-throughput study for many different clathrates. After creating a number of CEs for different clathrates one could then start to learn the ECIs given the chemical species involved. The latter suggestion is informed by paper III, in which trends in the ECIs emerged between Al or Ga-based clathrate. One could thus train a machine learning model to predict the ECIs for a given chemical composition, which only has to be accurate enough to restrict the search space.

In terms of ICET and its development many interesting and useful things can be added. To make high-throughput studies as mentioned above the creation of robust workflows would be invaluable. Although there is some degree of craftsmanship to creating a good CE, as we are becoming more experienced in using and creating CEs the procedures for doing routine tasks like training CEs is converging to algorithms that could very well be generalized into workflows. These workflows could do everything from generating suitable training structures to training CEs to running MC simulations on computer clusters and collecting the simulation data.

Acknowledgments

To my family, thank you for all your love and support.

To Emma, thank you for making our home such a warm and affectionate place.

To my friends, both local and those from up north, thank you for all the fun times.

To my supervisor Paul Erhart, I give you my sincere gratitude for giving me the opportunity to do this work. Your dedication and knowledge about physics and especially your enthusiasm to share it is greatly appreciated.

To my examiner, Göran Wahnström, thank you for your wisdom and leadership.

I would also like to thank Thomas Holm Rod and the whole data analysis & modelling group at ESS in Copenhagen for welcoming me to visit the group and see how you worked and developed software.

To my office mate, Erik Fransson, thank you for being such a good friend and rubber duck.

To William Armando Muñoz, Magnus Rahm, Erik Fransson and Paul Erhart, thank you for working with me to develop icet. It will be interesting to follow the development on the sideline from now on. Hopefully the bird will not be eaten by a cat anytime soon.

To Erik Fransson, Martin Gren and Göran Wahnström, thank you for letting me join in the WC project. It was both hard and tough, WC - that is.

To Adam Arvidsson and Anders Hellman, thank you for the expert counseling regarding zeolites.

To Joakim Brorsson and Anders Palmqvist, thank you for many interesting and helpful discussions regarding clathrates.

Finally, I am grateful for all current and former members of the Materials and Surface theory division that provide such a fun and exciting work environment.

Bibliography

[1] J. D. Hunter, *Matplotlib: A 2D graphics environment*, Computing in Science & Engineering **9**, 90 (2007).

[2] T. Williams, C. Kelley, and many others, *Gnuplot 4.6: an interactive plotting program*, `http://gnuplot.sourceforge.net/`, 2013.

[3] B. O. Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018.

[4] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, *The atomic simulation environment—a Python library for working with atoms*, Journal of Physics: Condensed Matter **29**, 273002 (2017).

[5] J. Walecka, *Introduction to Statistical Mechanics* (World Scientific, Singapore, 2011).

[6] D. Porter, K. Easterling, and M. Sherif, *Phase Transformations in Metals and Alloys, Third Edition (Revised Reprint)* (CRC Press, Boca Raton, Florida, 2009).

[7] M. Ångqvist, J. M. Rahm, L. Gharaee, and P. Erhart, *Structurally driven asymmetric miscibility in the phase diagram of W-Ti*, Phys. Rev. Materials **3**, 073605 (2019).

[8] J. M. Sanchez, F. Ducastelle, and D. Gratias, *Generalized cluster description of multicomponent systems*, Physica **128**, 334 (1984).

[9] A. van de Walle, *Multicomponent multisublattice alloys, nonconfigurational entropy and other additions to the Alloy Theoretic Automated Toolkit*, Calphad **33**, 266 (2009).

[10] E. Fransson, F. Eriksson, and P. Erhart, *Efficient construction and applications of higher-order force constant models*, 2019.

[11] L. J. Nelson, G. L. W. Hart, F. Zhou, and V. Ozoliņš, *Compressive sensing as a paradigm for building physics models*, Physical Review B **87**, 035125 (2013).

[12] L. J. Nelson, V. Ozoliņš, C. S. Reese, F. Zhou, and G. L. W. Hart, *Cluster expansion made easy with Bayesian compressive sensing*, Physical Review B **88**, 155105 (2013).

[13] E. Candes and M. Wakin, *An Introduction To Compressive Sampling*, IEEE Signal Processing Magazine **25**, 21 (2008).

[14] E. Candes and M. Wakin, *An Introduction To Compressive Sampling*, IEEE Signal Processing Magazine **25**, 21 (2008).

[15] H. Zou, *The Adaptive Lasso and Its Oracle Properties*, Journal of the American Statistical Association **101**, 1418 (2006).

[16] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **68**, 49 (2006).

[17] J. D. MacKay, *Bayesian nonlinear modeling for the prediction competition*, ASHRAE Transactions , (1994).

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research **12**, 2825 (2011).

[19] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, *API design for machine learning software: experiences from the scikit-learn project*, 2013.

[20] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of State Calculations by Fast Computing Machines*, The Journal of Chemical Physics **21**, 1087 (1953).

[21] B. Sadigh and P. Erhart, *Calculations of excess free energies of precipitates via direct thermodynamic integration across phase boundaries*, Physical Review B **86**, 134204 (2012).

[22] G. Ceder, *A derivation of the Ising model for the computation of phase diagrams*, Computational Materials Science **1**, 144 (1993).

[23] G. D. Garbulsky and G. Ceder, *Effect of lattice vibrations on the ordering tendencies in substitutional binary alloys*, Phys. Rev. B **49**, 6327 (1994).

[24] A. van de Walle and G. Ceder, *The effect of lattice vibrations on substitutional alloy thermodynamics*, Rev. Mod. Phys. **74**, 11 (2002).

[25] B. Sadigh, P. Erhart, A. Stukowski, A. Caro, E. Martinez, and L. Zepeda-Ruiz, *Scalable parallel Monte Carlo algorithm for atomistic simulations of precipitation in alloys*, Physical Review B **85**, 184203 (2012).

[26] P. Rogl, *Thermoelectrics Handbook* (CRC Press, Boca Raton, Florida, 2005), Chap. 32, pp. 1–24.

[27] A. V. Shevelkov and K. Kovnir, in *Zintl Phases*, No. 139 in *Structure and Bonding*, edited by T. F. Fässler (Springer, Heidelberg, Germany, 2011), pp. 97–142.

[28] T. F. T. Cerqueira, S. Pailhès, R. Debord, V. M. Giordano, R. Viennois, J. Shi, S. Botti, and M. A. L. Marques, *Prediction and Synthesis of a Non-Zintl Silicon Clathrate*, Chemistry of Materials **28**, 3711 (2016).

[29] G. S. Nolas, J. L. Cohn, G. A. Slack, and S. B. Schujman, *Semiconducting Ge clathrates: Promising candidates for thermoelectric applications*, Applied Physics Letters **73**, 178 (1998).

[30] E. S. Toberer, M. Christensen, B. B. Iversen, and G. J. Snyder, *High temperature thermoelectric efficiency in* $Ba_8Ga_{16}Ge_{30}$, Physical Review B **77**, 075203 (2008).

[31] B. B. Iversen, A. E. Palmqvist, D. E. Cox, G. S. Nolas, G. D. Stucky, N. P. Blake, and H. Metiu, *Why are Clathrates Good Candidates for Thermoelectric Materials?*, Journal of Solid State Chemistry **149**, 455 (2000).

[32] M. A. Avila, K. Suekuni, K. Umeo, H. Fukuoka, S. Yamanaka, and T. Takabatake, $Ba_8Ga_{16}Sn_{30}$ *with type-I clathrate structure: Drastic suppression of heat conduction*, Applied Physics Letters **92**, 041901 (2008).

[33] S. Christensen, M. S. Schmøkel, K. A. Borup, G. K. H. Madsen, G. J. McIntyre, S. C. Capelli, M. Christensen, and B. B. Iversen, *"Glass-like" thermal conductivity gradually induced in thermoelectric* $Sr_8Ga_{16}Ge_{30}$ *clathrate by off-centered guest atoms*, Journal of Applied Physics **119**, 185102 (2016).

[34] G. K. H. Madsen, K. Schwarz, P. Blaha, and D. J. Singh, *Electronic structure and transport in type-I and type-VIII clathrates containing strontium, barium, and europium*, Physical Review B **68**, 125212 (2003).

[35] T. Matsui, J. Furukawa, K. Tsukamoto, H. Tsuda, and K. Morii, *Structure and properties of $Ba_8Ga_{16}Ge_{30}$ clathrates by a novel synthesis method using CO gas reductive atmosphere*, Journal of Alloys and Compounds **391**, 284 (2005).

[36] N. P. Blake, S. Latturner, J. D. Bryan, G. D. Stucky, and H. Metiu, *Band structures and thermoelectric properties of the clathrates $Ba_8Ga_{16}Ge_{30}$, $Sr_8Ga_{16}Ge_{30}$, $Ba_8Ga_{16}Si_{30}$, and $Ba_8In_{16}Sn_{30}$*, The Journal of Chemical Physics **115**, 8060 (2001).

[37] M. Christensen, S. Johnsen, and B. B. Iversen, *Thermoelectric clathrates of type I*, Dalton Transactions **39**, 978 (2010).

[38] H. Schäfer, B. Eisenmann, and W. Müller, *Zintl Phases: Transitions between Metallic and Ionic Bonding*, Angew. Chem., Int. Ed. Engl. **12**, 694 (1973).

[39] K. A. Kovnir and A. V. Shevelkov, *Semiconducting clathrates: synthesis, structure and properties*, Russian Chemical Reviews **73**, 923 (2004).

[40] J. H. Roudebush, C. de la Cruz, B. C. Chakoumakos, and S. M. Kauzlarich, *Neutron Diffraction Study of the Type I Clathrate $Ba_8Al_xSi_{46-x}$: Site Occupancies, Cage Volumes, and the Interaction between the Guest and the Host Framework*, Inorganic Chemistry **51**, 1805 (2012).

[41] M. Ångqvist and P. Erhart, *Understanding Chemical Ordering in Intermetallic Clathrates from Atomic Scale Simulations*, Chemistry of Materials **29**, 7554 (2017).

[42] M. Ångqvist, D. O. Lindroth, and P. Erhart, *Optimization of the Thermoelectric Power Factor: Coupling between Chemical Order and Transport Properties*, Chem. Mater. **28**, 6877 (2016).

[43] M. Troppenz, S. Rigamonti, and C. Draxl, *Predicting Ground-State Configurations and Electronic Properties of the Thermoelectric Clathrates $Ba_8Al_xSi_{46-x}$ and $Sr_8Al_xSi_{46-x}$*, Chem. Mater. **29**, 2414 (2017).

[44] *Binary Alloy Phase Diagram*, edited by T. Massalski (ASM International, Materials Park, Ohio, ADDRESS, 1990).

[45] R. Najafabadi, D. J. Srolovitz, E. Ma, and M. Atzmon, *Thermodynamic properties of metastable Ag-Cu alloys*, Journal of Applied Physics **74**, 3144 (1993).

[46] P. L. Williams, Y. Mishin, and J. C. Hamilton, *An embedded-atom potential for the Cu–Ag system*, Modelling and Simulation in Materials Science and Engineering **14**, 817 (2006).

[47] R. R. Hultgren and P. D. Desai, *Selected Thermodynamic Values and Phase Diagrams for Copper and Some of Its Binary Alloys* (International Copper Research Association, New York, 1971).

[48] J. Jiang, C. Gente, and R. Bormann, *Mechanical alloying in the Fe–Cu system*, Materials Science and Engineering: A **242**, 268 (1998).

[49] E. M. Lopasso, M. Caro, A. Caro, and P. E. A. Turchi, *Phase diagram of an empirical potential: The case of Fe-Cu*, Physical Review B **68**, 214205 (2003).

[50] V. Ozoliņš, C. Wolverton, and A. Zunger, *Cu-Au, Ag-Au, Cu-Ag, and Ni-Au intermetallics: First-principles study of temperature-composition phase diagrams and structures*, Physical Review B **57**, 6427 (1998).

[51] M. Muzyk, D. Nguyen-Manh, K. J. Kurzydłowski, N. L. Baluc, and S. L. Dudarev, *Phase stability, point defects, and elastic properties of W-V and W-Ta alloys*, Physical Review B **84**, 104115 (2011).

[52] *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology*, edited by O. Madelung (Springer-Verlag, Heidelberg, ADDRESS, 1996), Vol. 5.

[53] P. Fredriksson and B. Sundman, *A thermodynamic assessment of the Fe-Pt system*, Calphad **25**, 535 (2001).

[54] J. Kim, Y. Koo, and B.-J. Lee, *Modified embedded-atom method interatomic potential for the Fe–Pt alloy system*, Journal of Materials Research **21**, 199–208 (2006).

[55] I. Ansara, B. Sundman, and P. Willemin, *Thermodynamic modeling of ordered phases in the Ni-Al system*, Acta Metallurgica **36**, 977 (1988).

[56] G. P. Pun and Y. Mishin, *Development of an interatomic potential for the Ni-Al system*, Philosophical Magazine **89**, 3245 (2009).

[57] A. Caro, M. Caro, E. M. Lopasso, and D. A. Crowson, *Implications of ab initio energetics on the thermodynamics of Fe-Cr alloys*, Applied Physics Letters **89**, 121902 (2006).

[58] G. Bonny, R. C. Pasianot, and L. Malerba, *Interatomic potentials for alloys: Fitting concentration dependent properties*, Philosophical Magazine **89**, 711 (2009).

[59] G. J. Ackland, *Magnetically Induced Immiscibility in the Ising Model of FeCr Stainless Steel*, Physical Review Letters **97**, 015502 (2006).

[60] W. Petry, A. Heiming, J. Trampenau, M. Alba, C. Herzig, H. R. Schober, and G. Vogl, *Phonon dispersion of the bcc phase of group-IV metals. I. bcc titanium*, Physical Review B **43**, 10933 (1991).

[61] K. Einarsdotter, B. Sadigh, G. Grimvall, and V. Ozoliņš, *Phonon Instabilities in fcc and bcc Tungsten*, Physical Review Letters **79**, 2073 (1997).

[62] V. Ozolins, *First-Principles Calculations of Free Energies of Unstable Phases: The Case of fcc W*, Physical Review Letters **102**, 065702 (2009).

[63] L. Kaufman and H. Nesor, *Calculation of Superalloy Phase Diagrams: Part IV*, Metallurgical Transactions **6A**, 2123 (1975).

[64] J. L. Murray, *The Ti-W (Titanium-Tungsten) System*, Bulletin of Alloy Phase Diagrams **2**, 192 (1981).

[65] S. K. Lee and D. N. Lee, *Calculation of phase diagrams using partial phase diagram data*, CALPHAD **10**, 61 (1986).

[66] S. Jönsson, *Reevaluation of the Ti-W system and prediction of the Ti-W-N phase diagram*, Zeitschrift für Metallkunde **87**, 784 (1996).

[67] A. S. Kurlov and A. I. Gusev, *Introduction*, in *Tungsten Carbides: Structure, Properties and Application in Hardmetals* (Springer International Publishing, Cham, 2013), p. 1.

[68] Y. Xiao, J.-Y. Hwang, and Y.-K. Sun, *Transition metal carbide-based materials: synthesis and applications in electrochemical energy storage*, J. Mater. Chem. A **4**, 10379 (2016).

[69] W. S. Williams, *Physics of transition metal carbides*, Mat. Sci Eng. A-Struct **105**, 1 (1988).

[70] H. O. Pierson, in *Handbook of Refractory Carbides and Nitrides*, edited by H. O. Pierson (William Andrew Publishing, Westwood, NJ, 1996), p. 8.

[71] S. Norgren, J. García, A. Blomqvist, and L. Yin, *Trends in the P/M hard metal industry*, International Journal of Refractory Metals and Hard Materials **48**, 31 (2015).

[72] A. S. Kurlov and A. I. Gusev, *Tungsten Carbides and W-C Phase Diagram*, Inorganic Materials **42**, 121 (2006).

[73] S. Lay and J.-M. Missiaen, in *Comprehensive Hard Materials*, edited by V. K. Sarin (Elsevier, Oxford, 2014), p. 91.

[74] V. Bounhoure, S. Lay, M. Loubradou, and J.-M. Missiaen, *Special WC/Co Orientation Relationships at Basal Facets of WC Grains in WC–Co Alloys*, Journal of Materials Science **43**, 892 (2008).

[75] X. Liu, X. Song, H. Wang, X. Liu, F. Tang, and H. Lu, *Complexions in WC-Co Cemented Carbides*, Acta Materialia **149**, 164 (2018).

[76] I. Konyashin, A. Sologubenko, T. Weirich, and B. Ries, *Complexion at WC-Co grain boundaries of cemented carbides*, Materials Letters **187**, 7 (2017).

[77] S. Lay, C. H. Allibert, M. Christensen, and G. Wahnström, *Morphology of WC Grains in WC–Co Alloys*, Materials Science and Engineering: A **486**, 253 (2008).

[78] A. Sivkov, A. Pak, and I. Rakhmatullin, *Production of ultrafine tungsten carbide in a discharge plasma jet.*, Nanotechnol Russia **9**, 682 (2014).

[79] C. Baerlocher and L. McCusker, *Database of Zeolite Structures*, `http://www.iza-structure.org/databases/`.

[80] M. Jeffroy, C. Nieto-Draghi, and A. Boutin, *New Molecular Simulation Method To Determine Both Aluminum and Cation Location in Cationic Zeolites*, Chemistry of Materials **29**, 513 (2017).

[81] B. C. Knott, C. T. Nimlos, D. J. Robichaud, M. R. Nimlos, S. Kim, and R. Gounder, *Consideration of the Aluminum Distribution in Zeolites in Theoretical and Experimental Catalysis Research*, ACS Catalysis **8**, 770 (2018).

[82] W. Loewenstein, *The distribution of aluminum in the tetrahedra of silicates and aluminates*, American Mineralogist **39**, 92 (1954).

[83] R. G. Bell, R. A. Jackson, and C. R. A. Catlow, *Löwenstein's rule in zeolite A: A computational study*, Zeolites **12**, 870 (1992).

[84] R. E. Fletcher, S. Ling, and B. Slater, *Violations of Löwenstein's rule in zeolites*, Chemical Science **8**, 7483 (2017).

[85] F. Eriksson, E. Fransson, and P. Erhart, *The Hiphive Package for the Extraction of High-Order Force Constants by Machine Learning*, Adv. Theory Simul. **0**, 1800184 (2019).

[86] P. A. Korzhavyi, L. V. Pourovskii, H. W. Hugosson, A. V. Ruban, and B. Johansson, *Ab Initio Study of Phase Equilibria in* TiC_x, Phys. Rev. Lett. **88**, 015505 (2001).

[87] K. F. Garrity, *Combined cluster and atomic displacement expansion for solid solutions and magnetism*, Phys. Rev. B **99**, 174108 (2019).